


```
XX PI Gladyshev VN, Wootton JC, Hatfield DL, Jeang K;
XX DR WPt: 2000-013045/01.
XX N-PSDB: AAZ34464.
XX PT New isolated selenoprotein polypeptides, used to develop products for
PT detecting susceptibility to or treating cancers e.g. prostate cancers
PR -
PS Claim 1; Page 55; 67pp; English.
XX XX
XX The present sequence represents a human 15 kDa selenoprotein that
CC is differentially expressed in cancer cells, such as cancer cells.
CC It includes a selenocysteine residue at position 93 that is encoded
CC by a TGA codon in the cDNA (see AAZ34464). Polymorphisms in the
CC 15 kDa selenoprotein gene are associated with susceptibility to
CC cancer. Claimed methods of detecting the presence, or of
CC quantifying the level of expression, of the 15 kDa selenoprotein
CC involve contacting a sample with a specific binding agent, such as
CC a polyclonal antibody or monoclonal antibody, that specifically
CC binds to the 15 kDa selenoprotein. A claimed method for dietary
CC regulation comprises detecting an abnormally low expression of a 15
CC kDa selenoprotein in a subject and, if the level is below normal,
CC enhancing the level by providing additional selenium in the diet.
CC The susceptibility to cancer of a subject having an increased
CC predetermined genetic susceptibility to cancer is reduced by
CC administering the 15 kDa selenoprotein or by overexpressing the
CC selenoprotein using gene therapy.
XX
XX Sequence    162 AA:
S0
Query Match                99.8%; Score 829; DB 21; Length 162;
Best Local Similarity     100.0%; Pred. No. 8.8e-88;
Matches   162; Conservative      0; Mismatches      0; Gaps      0;
OY       1 MAAGSGCLVPAFRGRLLIATYLQAVSAFGAFSSBACHRELGFSSNLCSDDLGGFPL 60
Dd        | |||||
Db         1 maagsgclypafrgrlllatylqavsafgaefssaeacrelgfsanlilcsscdllggfnl 60
OY       61 IQLDPDCRGCCEEEHOFETKKLYACALIEVCXKIGRPVOVAFYRSDDPKLFRGLQTIKY 120
Dd        | |||||
Db         61 lqldpdcrgccgeeehfetkrlkyagailievngxkligrfpvqatyrddpkflrfgltiky 120
OY       121 VRGSDPVTLKLDDNGNIAEELSLKMNTDSVEFTSEKIERTI 162
Dd        | |||||
Db        121 vrgsdpvtlklldngniaeelslkmntdsveeftsekiertl 162
RESULT          2
AAV32113
ID            AAY32113 standard; Protein: 162 AA.
XX
XX           AAY32113;
XX CC
DX             01-FEB-2000 (first entry)
XX DT
XX DE              Mouse 15 kDa selenoprotein.
XX XX
KW Selenoproteins; selenium; differential expression; tumour; mouse;
KW prostate cancer; diagnosis; polymorphism.
XX KW
OS Mus musculus.
XX OS
PH Key                      Location/Qualifiers
FH Modified-site             93 OTHER
FT //label= "OTHER
FX /note= "selenocysteine"
PN MO9951637-A1.
XX PD
PD 14-OCT-1999.
```

XX	06-APR-1999;	99WO-US07560.
PE		
XX	06-APR-1998;	98US-0080850.
PR		
XX	(USSH) US DEPT HEALTH & HUMAN SERVICES.	
PA		
XX	Gladyshev VN, Wootton JC, Hatfield DL, Jeang K;	
PI		
XX	WPI: 2000-013045/01.	
DR	N-PSDB: AA234469.	
XX		
PT	New isolated selenoprotein polypeptides, used to develop products for	
PT	detecting susceptibility to or treating cancers e.g. prostate cancers	
PT		
XX		
XX	Claim 1; Page 60; 67pp; English.	
PS		
XX		
CC	The present sequence represents a mouse 15 kDa selenoprotein that	
CC	is differentially expressed in cancer cells, such as cancer cells.	
CC	It includes a selenocysteine residue at position 93 that is encoded	
CC	by a TGA codon in the cDNA (see AA234468). Polymorphisms in the	
CC	human 15 kDa selenoprotein gene are associated with susceptibility	
CC	to cancer. Claimed methods of detecting the presence, or of	
CC	quantifying the level of expression, of the 15 kDa selenoprotein	
CC	involve contacting a sample with a specific binding agent, such as	
CC	a polyclonal antibody or monoclonal antibody, that specifically	
CC	binds to the 15 kDa selenoprotein. A claimed method for dietary	
CC	regulation comprises detecting an abnormally low expression of a	
CC	mammalian 15 kDa selenoprotein in a subject and, if the level is	
CC	below normal, enhancing the level by providing additional selenium	
CC	in the diet. The susceptibility to cancer of a subject having an	
CC	increased predetermined genetic susceptibility to cancer is reduced	
CC	by administering a mammalian 15 kDa selenoprotein or by	
CC	overexpressing the selenoprotein using gene therapy.	
CC		
XX	Sequence 162 AA:	
SO		
	Query Match 90.7%; Score 754; DB 21; Length 162;	
	Best Local Similarity 92.6%; Pred. No. 4.3e-79;	
	Matches 150; Conservative 1; Mismatches 11; Indels 0; Gaps 0	
QY	1 MAAGSGCLVPAFGKRLLIATVLAQVSAFGAFSSBACRELGFSSMLCCSDLLGQFNL 60	
DB	1 maaqgggwlrlpalgjlrlilatafgvsalgaefaseacrelgfssnllcscdllyqfql 60	
QY	61 LQLDDCGACCOEEOEFKKRLYGAGLLEVCXKIKGRFGVQAVFASDKPKLFRGIQIR 120	
DB	61 lpldvcrcgcqeeaeqfekkllygaqalleevgxxklygrfpvqafvrsdkpklyfrglqiky 120	

KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
 XX Homo sapiens.
 OS
 XX MO9842738-A1.
 XX
 PD 01-OCT-1998.
 XX
 PF 19-MAR-1998; 98WO-US05311.
 XX
 PR 30-MAY-1997; 97US-0050937.
 PR 21-MAR-1997; 97US-0041276.
 PR 21-MAR-1997; 97US-0041277.
 PR 21-MAR-1997; 97US-0041281.
 PR 21-MAR-1997; 97US-0042344.
 PR 30-MAY-1997; 97US-0048069.
 PR 30-MAY-1997; 97US-0048094.
 PR 30-MAY-1997; 97US-0048095.
 PR 30-MAY-1997; 97US-0048096.
 PR 30-MAY-1997; 97US-0048099.
 PR 30-MAY-1997; 97US-0048131.
 PR 30-MAY-1997; 97US-0048135.
 PR 30-MAY-1997; 97US-0048160.
 PR 30-MAY-1997; 97US-0048186.
 PR 30-MAY-1997; 97US-0048187.
 PR 30-MAY-1997; 97US-0048188.
 PR 30-MAY-1997; 97US-0048350.
 PR 30-MAY-1997; 97US-0048351.
 PR 30-MAY-1997; 97US-0048352.
 PR 30-MAY-1997; 97US-0048355.
 PR 05-AUG-1997; 97US-0054804.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Brewer LA, Duan R, Ebner R, Ferrle AM, Florence KA;
 PI Greene JM, Hu JS, Lafleur DM, Moore PA, Ni J, Olsen HS;
 PI Rosen CA, Ruben SM, Shi Y, Young P;
 XX
 DR N-PSDB; AAX00680.
 DR
 XX WPI: 1999-070066/06.
 PT New isolated human genes and the secreted polypeptides they encode -
 PT useful for diagnosis and treatment of e.g. cancers, neurological
 PT disorders, immune diseases, inflammation or blood disorders
 PS
 XX Claim 11; Page 310; 385pp; English.
 PS
 CC This sequence represents a secreted human protein encoded by the gene
 CC clone detailed in the descriptor line. The gene can be used to generate
 CC fusion proteins by linking to the gene to a human immunoglobulin Fc
 CC portion (e.g. AAX00602) for increasing the stability of the fused
 CC protein as compared to the human protein only.
 CC The invention relates to 87 novel genes and their fragments (nucleic
 CC acid sequences: AAX00611-X00724; amino acid sequences AAW67807-W68004)
 CC which are useful for preventing, treating or ameliorating medical
 CC conditions e.g. by protein or gene therapy. Also, pathological
 CC conditions can be diagnosed by determining the amount of the new
 CC polypeptides in a sample or by determining the presence of mutations in
 CC the new polynucleotides. Specific uses are described for each of the 87
 CC polynucleotides, based on which tissues they are most highly expressed in
 CC (see AAX00611 for described uses).
 CC
 XX
 SQ Sequence 92 AA;
 Query Match 57.0%; Score 474; DB 20; Length 92;
 Best Local Similarity 98.9%; Pred. No. 5.2e-47;
 Matches 91; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Oy 1 MAAGSGCLVPAFGKRLILATVLAQVAFGAFFSSSEACRELGSSNLLCSCDLGQFNL 60
 Db 1 maagsgclvpafigrlllatvylaqvafgaaffssseacrelgssnllcscdlgqfnl 60

Oy 61 IQLDPDRCGCCQEAQFETKRLYAGAILVECG 92
 Db 61 Iqldpdrcgcccqeaqfektlyagaillevcg 92
 RESULT 4
 ID AAY65395 standard; Protein; 95 AA.
 XX AAY65395;
 AC
 XX
 DT 01-FEB-2000 (first entry)
 XX
 DE Human 5' EST related polypeptide SEQ ID NO:1556.
 XX
 KW Human: 5' EST; expressed sequence tag; secreted protein; diagnosis;
 KW gene therapy; chromosome mapping; upstream regulatory sequence;
 KW forensic; location; development; protein synthesis; stability;
 KW regulation; identification.
 XX
 OS Homo sapiens.
 XX
 PN WO9953051-A2.
 PD 21-OCT-1999.
 XX
 PF 09-APR-1999; 99WO-IB00712.
 PR 09-APR-1998; 98US-0057719.
 PR 28-APR-1998; 98US-0069047.
 XX
 PA (GEST) GENSET.
 XX
 PI Dumas Milne Edwards J, Duclert A, Giordano J;
 XX
 DR WPI: 2000-038446/03.
 DR N-PSDB; AAZ43009.
 XX
 PT Novel secreted protein 5' expressed sequence tag sequences used in
 PT diagnostic, forensic, gene therapy, and chromosome mapping procedures
 PS
 XX Claim 3; Page 818; 837pp; English.
 PS
 CC AAZ42265 to AAZ43075 represent novel 5' expressed sequence tag (EST)
 CC sequences, corresponding to human secreted proteins. AAY64651 to
 CC AAY65438 represent the EST-related proteins corresponding to AAZ42265 to
 CC AAZ43052. The 5' ESTs can be used for producing secreted human gene
 CC products. They can be used to identify and isolate 5' untranslated
 CC regions (UTRs) and upstream regulatory regions which control the
 CC location, development stage, rate, and quantity of protein synthesis, as
 CC well as stability of mRNA. The ESTs are also useful as probes for
 CC chromosome mapping, and to obtain full length cDNA clones. The ESTs can
 CC also be used in forensic procedures to identify individuals, or in
 CC diagnostic procedures to identify individuals having genetic diseases
 CC resulting from abnormal gene expression. The products may also be used in
 CC gene therapy protocols. The nucleic acids encoding signal peptides can be
 CC used for directing extracellular secretion of a polypeptide or the
 CC insertion of a polypeptide into a membrane, or importing a polypeptide
 CC into a cell. The proteins encoded by the EST sequences may be useful in
 CC treating a variety of human conditions. Secreted proteins have
 CC therapeutic value, and the identification of new secreted proteins is
 CC valuable. AAZ42249 to AAZ42264 and AAY64644 to AAY64650 represent
 CC sequences used in the exemplification of the present invention.
 CC
 XX
 SQ Sequence 95 AA;
 Query Match 57.0%; Score 474; DB 21; Length 95;
 Best Local Similarity 98.9%; Pred. No. 5.4e-47;
 Matches 91; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Oy 1 MAAGSGCLVPAFGKRLILATVLAQVAFGAFFSSSEACRELGSSNLLCSCDLGQFNL 60

|||||
Db 4 maagpsgclvpaflrlllatvqvasaifaetssaeacrelgfsnllcscdlldlqfnnl 63
QY 61 LQLDPCRGCCQEEAFQETKKLYAGAILLEVCG 92
|
Db 64 lqldpcrcgcgcgaefetkkllyagallevcg 95

RESULT 5
AAB57132
ID AAB57132 standard; Protein; 124 AA.
XX
AC AAB57132;
XX

DT 13-MAR-2001 (first entry)
XX

DE Human prostate cancer antigen protein sequence SEQ ID NO:1710.
XX

KM Human: prostate cancer; prostate cancer antigen; detection; diagnosis;
neuroprotective; cytosolic; cardioactive; immunomodulatory; muscular;
KW vulnary; gastrointestinal; nephrotoxic; antineoplastic; gynaecological;
KW antibacterial; gene therapy; neural; immune; reproductive; renal;
KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
KW wound; infectious disease.
XX

OS Homo sapiens.
XX

PN WO20005174-A1.
XX

PD 21-SEP-2000.
XX

PF 08-MAR-2000; 2000MO-US05988.
XX

PR 12-MAR-1999; 9905-0124270.
XX

PA (HUMA-) HUMAN GENOME SCI INC.
(ROSE/) ROSEN C A.
XX

PI Rosen CA, Ruben SM;
XX

DR WPI: 2000-587513/55.
XX

DR N-PSDB; AAF16335.
XX

PT Prostate cancer associated gene sequences, referred to as prostate
cancer antigens, useful for treatment, prevention, and diagnosis of
PT disorders such as prostate cancer -
XX

PS Claim 11; Page 2183-2184; 2338pp; English.
XX

CC AAF15566 to AAF16505 encode the human prostate cancer associated
proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
CC The prostate cancer antigens can have neuroprotective, cytosolic,
CC cardioactive, immunomodulatory, muscular, vulnary, gastrointestinal,
CC nephrotoxic, antineoplastic, gynaecological and antibacterial activities,
CC and can be used in gene therapy. The prostate cancer antigen
CC polynucleotides may be used for detection of prostate cancer, chromosome
identification, as chromosome markers, and for numerous other diagnostic
CC or research purposes. The prostate cancer antigens may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to
CC AAB57303 represent sequences used in the exemplification of the present
CC invention.
XX

XX Sequence 124 AA;
SQ

Query Match 57.0%; Score 474; DB 21; Length 124;
Best Local Similarity 98.9%; Pred. No. 7.8e-47;

Matches 91; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAAGPSGCLVPAFGKRLLATVQAVSAFGAEEFSSAEACRELGFSSNLLCSCDLIDLGQFNL 60
|
|

Db 33 maagpsgclvpaflrlllatvqvasaifaetssaeacrelgfsnllcscdlldlqfnnl 92
QY 61 LQLDPCRGCCQEEAFQETKKLYAGAILLEVCG 92
|
Db 93 lqldpcrcgcgcgaefetkkllyagallevcg 124

RESULT 6
ABB63319
ID ABB63319 standard; Protein; 142 AA.
XX
AC ABB63319;
XX

DT 26-MAR-2002 (first entry)
XX

DE Drosophila melanogaster polypeptide SEQ ID NO 16749.
XX

KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX

OS Drosophila melanogaster.
XX

PN WO200171042-A2.
XX

PD 27-SEP-2001.
XX

PF 23-MAR-2001; 2001MO-US09231.
XX

PR 23-MAR-2000; 2000US-191637P.
XX

PR 11-JUL-2000; 2000US-0614150.
XX

PA (PEKE) PE CORP NY.
XX

PI Venter JC, Adams M, Li PMD, Myers EW;
XX

DR WPI: 2001-656860/75.
XX

DR N-PSDB; ABL07422.
XX

PT New isolated nucleic acid detection reagent for detecting 1000 or more
genes from Drosophila and for elucidating cell signalling and cell-cell
interactions -
XX

PS Disclosure; SEQ ID NO 16749; 21pp + Sequence Listing; English.
XX

CC The invention relates to an isolated nucleic acid detection reagent
capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
sequences (AB101840-AB16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pct_sequences.
XX

XX Sequence 142 AA;
SQ

Query Match 37.4%; Score 311; DB 22; Length 142;
Best Local Similarity 51.8%; Pred. No. 7.1e-28;

Matches 59; Conservative 17; Mismatches 38; Indels 0; Gaps 0;

QY 48 LCSSCDLILGQFNLQLDPCRGCCQEEAFQETKKLYAGAILLEVCGXKLGFPQVAFVRS 107
|
|
Db 1 mscsccklddfgldtlkpgckqccctldqgpaqctlykallewtckfrfypqatqigs 60

QY 108 DKRPLFGLQIKYRGSDPVKLLDDNGNTAEELSTKMNWDVSEPLSEKLER 161
|
|
Db 61 grpafkfnldikyrgldpvklldaagkvgetlstkntdveeffechnlak 114

RESULT 7

AAB75288
ID AAB75288 standard; Protein; 60 AA.
XX
AC AAB75288;
XX
DT 03-APR-2001 (first entry)
XX
DE Gene 1 human secreted protein homologous amino acid sequence #107.
XX
KW Human; immunosuppressive; antiarthritic; antirheumatic; neurotropic;
KW antiproliferative; cytosstatic; cardiant; vasotropic; cerebroprotective;
KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
KW vulnery; autoimmune disease; hyperproliferative disorder; cancer;
KW cardiovascular disorder; cerebrovascular disorder; infection;
KW nervous system disorder; ocular disorder; chemotaxis; food additive;
secreted protein.
XX
OS Homo sapiens.
XX
PN WO200077021-A1.
XX
PD 21-DEC-2000.
XX
PF 01-JUN-2000; 2000MO-US15135.
XX
PR 11-JUN-1999; 9905-0138632.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX (ROSE/) ROSEN C A.
PI Rosen CA, Ruben SM, Komatsoulis GA;
XX WPI: 2001-071257/08.
DR
XX
PT Nucleic acid molecules encoding human secreted proteins, used in
PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
PT Parkinson's diseases and cancers -
XX
PS Disclosure; Page 10; 530pp; English.
XX
CC This invention relates to polynucleotide sequences AAF63789 - AAF63836
CC which encode human secreted proteins AAB75260 - AAB75287. Included in the
CC invention are protein sequences AAB75288 - AAB75341 which are fragments
CC of the secreted proteins and amino acid sequences with which these
CC fragments share homology. Examples of the activities of the proteins and
CC polynucleotides and the activities of their agonists and antagonists
CC include, immunosuppressive; antiarthritic; antirheumatic;
CC antiproliferative; cytosstatic; cardiant; vasotropic; cerebroprotective;
CC neurotropic; neuroprotective; antibacterial; virucide; fungicide;
CC ophthalmological; and vulnery activity. The protein and polynucleotide
CC sequences, their agonists and antagonists may be useful for treating,
CC preventing and diagnosing diseases and disorders such as autoimmune
CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders
CC e.g. neoplasms of the breast or liver, cardiovascular disorders
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC anglogenesis, nervous system disorders e.g. Alzheimer's disease,
CC infections caused by bacteria, viruses and fungi and ocular disorders
CC e.g. corneal infection. The polypeptides can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities. Included in the invention are
CC oligonucleotides AAF63780 - AAF63788 and peptide AAB75239 which are used
CC in the identification and characterisation of the DNA and protein
CC sequences of the invention.
XX
SQ Sequence 60 AA:

Query Match 36.3%; Score 302; DB 22; Length 60;
Best Local Similarity 100.0%; Pred. No. 2.4e-27;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 103 AAFRSDKPKLFRLGIQIRYRGSDPVILKLDNDNGNIAEELSILKWNDSVEEFLSEKLERI 162
DB 1 afvrsdkpkifrglqikvyrsgdpvilkldndngniaeelsilkwntdsveeflsekleri 60
RESULT 8
AAB75289
ID AAB75289 standard; Protein; 60 AA.
XX
AC AAB75289;
XX
DT 03-APR-2001 (first entry)
XX
DE Human secreted protein sequence encoded by gene 1 SEQ ID NO:108.
XX
KW Human; immunosuppressive; antiarthritic; antirheumatic; neurotropic;
KW antiproliferative; cytosstatic; cardiant; vasotropic; cerebroprotective;
KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
KW vulnery; autoimmune disease; hyperproliferative disorder; cancer;
KW cardiovascular disorder; cerebrovascular disorder; infection;
KW nervous system disorder; ocular disorder; chemotaxis; food additive;
secreted protein.
XX
OS Homo sapiens.
XX
PN WO200077021-A1.
XX
PD 21-DEC-2000.
XX
PF 01-JUN-2000; 2000MO-US15135.
XX
PR 11-JUN-1999; 9905-0138632.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX (ROSE/) ROSEN C A.
PI Rosen CA, Ruben SM, Komatsoulis GA;
XX WPI: 2001-071257/08.
DR N-PSDB; AAF63789.
XX
PT Nucleic acid molecules encoding human secreted proteins, used in
PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
PT Parkinson's diseases and cancers -
XX
PS Disclosure; Page 10; 530pp; English.
XX
CC This invention relates to polynucleotide sequences AAF63789 - AAF63836
CC which encode human secreted proteins AAB75260 - AAB75287. Included in the
CC invention are protein sequences AAB75288 - AAB75341 which are fragments
CC of the secreted proteins and amino acid sequences with which these
CC fragments share homology. Examples of the activities of the proteins and
CC polynucleotides and the activities of their agonists and antagonists
CC include, immunosuppressive; antiarthritic; antirheumatic;
CC antiproliferative; cytosstatic; cardiant; vasotropic; cerebroprotective;
CC neurotropic; neuroprotective; antibacterial; virucide; fungicide;
CC ophthalmological; and vulnery activity. The protein and polynucleotide
CC sequences, their agonists and antagonists may be useful for treating,
CC preventing and diagnosing diseases and disorders such as autoimmune
CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders
CC e.g. neoplasms of the breast or liver, cardiovascular disorders
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC anglogenesis, nervous system disorders e.g. Alzheimer's disease,
CC infections caused by bacteria, viruses and fungi and ocular disorders
CC e.g. corneal infection. The polypeptides can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities. Included in the invention are
CC oligonucleotides AAF63780 - AAF63788 and peptide AAB75239 which are used
CC in the identification and characterisation of the DNA and protein

CC sequences of the invention.
XX Sequence 60 AA:
SQ

Query Match 36.3%; Score 302; DB 22; Length 60;
Best Local Similarity 100.0%; Pred. No. 2.4e-27;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 103 AFVRSDEKRLRGIOIKYVSGSDPYLKLDDNGNIAEELSLIKMNTDSEVEFLSEKLERI 162
Db 1 afvrsdkpklrtgldikyvrgsdpylklldngnaeelslkwntdsveeflsekleri 60
|||||

RESULT 9
AAG08634
ID AAG08634 standard; Protein: 149 AA.
XX
AC AAG08634;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 6249.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX
OS Arabidopsis thaliana.
XX
PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123160.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136302.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143942.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 22-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.

PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.
PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154033.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155133.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155489.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159325.
PR	18-OCT-1999;	99US-0159584.
PR	18-OCT-1999;	99US-0159638.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.
Query Match		
27.3%;		
Score 226.5; DB 21; Length 149;		
Beet Local Similarity 32.9%;		
Pred. No. 4.6e-18;		

Matches	48;	Conservative	37;	Mismatches	54;	Indels	7;	Gaps	4;
Qy	17	LILATVQLQVSAFACGFEFSSEACRELFQSSNLLCSCDGLGQP-NLLQLDPDRCGCCQEEA	75						
Db	4	lnlaastiske-----qlskcededigfsgjalcsdchslseykkgdelvsdclkcacads	59						
Qy	76	QEFETKKL-VAGAILLEVCGKIGRFPQVOAVFRSDKPKLFIRGLQIRYVRGSDPVLKLLDDN	134						
Db	60	edsmkskvtysgsllewcmmklvfyepvgffieekek-fpsvkvgylfinsppklmlded	118						
Qy	135	GNIAERLSILKWNNTDSVEEFLSEKLE	160						
Db	119	gehkesiridnwkrehllqymrekvk	144						
RESULT 10									
AACG08633									
ID	AACG08633 standard; Protein; 155 AA.								
XX	AACG08633;								
XX	AC								
XX	DT								
XX	17-OCT-2000 (first entry)								
XX	Arabidopsis thaliana protein fragment SEQ ID NO: 6248.								
XX	Protein identification: signal transduction pathway; metabolic pathway;								
XX	hybridisation assay; genetic mapping; gene expression control; promoter;								
XX	termination sequence.								
XX	Arabidopsis thaliana.								
XX	EP1033405-A2.								
XX	06-SEP-2000.								
PF	25-FEB-2000; 2000EP-0301439.								
XX									
PR	25-FEB-1999;	990S-0121825.							
PR	05-MAR-1999;	990S-0123180.							
PR	09-MAR-1999;	990S-0123548.							
PR	23-MAR-1999;	990S-0125788.							
PR	25-MAR-1999;	990S-0126264.							
PR	29-MAR-1999;	990S-0126785.							
PR	01-APR-1999;	990S-0127462.							
PR	06-APR-1999;	990S-0128234.							
PR	08-APR-1999;	990S-0128714.							
PR	16-APR-1999;	990S-0129845.							
PR	19-APR-1999;	990S-0130077.							
PR	21-APR-1999;	990S-0130449.							
PR	23-APR-1999;	990S-0130510.							
PR	23-APR-1999;	990S-0130891.							
PR	28-APR-1999;	990S-0131449.							
PR	30-APR-1999;	990S-0132048.							
PR	30-APR-1999;	990S-0132407.							
PR	04-MAY-1999;	990S-0132484.							
PR	05-MAY-1999;	990S-0132485.							
PR	06-MAY-1999;	990S-0132486.							
PR	06-MAY-1999;	990S-0132487.							
PR	07-MAY-1999;	990S-0132863.							
PR	11-MAY-1999;	990S-0134256.							
PR	14-MAY-1999;	990S-0134218.							
PR	14-MAY-1999;	990S-0134219.							
PR	14-MAY-1999;	990S-0134221.							
PR	14-MAY-1999;	990S-0134370.							

PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139819.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140891.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145813.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.

PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150586.
PR 27-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 26-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156536.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 23-OCT-1999; 99US-0160989.
PR 23-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match

27.38; Score 226.5; DB 21; Length 155;

Accession	Gene	Protein	Length	Score	E-value	Identity	Similarity	Positives	Negatives	Gaps
Best Local Similarity	32.9%	Pred. No. 4.8e-18;								
Matches 48;	Conservative 37;	Mismatches 54;	Indels 7;	Gaps 4;						
OY	17	ILILAVILQAVSAFGAEFSEACREIFGFSNNLCCSODLIGOF-NILQLDPDRCGCCQERA	75							
Db	10	lmlastlsake---qlstkecedigfsglaldscdchalseyvkqdelvsdcikccads	65							
OY	76	QFETKRL-YAGAILVEVCCKLGRFOVOAFVRSDPKLFRLGIQIKVRGSDPVLKLLDNN	134							
Db	66	edsmskvysgalllevcmnkrlvfyelvgfleecek-fpskvqyiftnspkllmided	124							
OY	135	GNIAEELSLKWNITDSVEEFLSEKIE	160							
Db	125	gekhesiridnwkrehllqymrekvk	150							
RESULT	11									
AAAG43258										
ID	AAAG43258	standard; Protein; 155 AA.								
XX	AAAG43258;									
XX	18-OCT-2000	(first entry)								
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 54048.									
KW	Protein identification; signal transduction pathway; metabolic pathway;									
KW	hybridisation assay; genetic mapping; gene expression control; promoter;									
KW	termination sequence.									
XX	Arabidopsis thaliana.									
OS	EP1033405-A2.									
PN	06-SEP-2000.									
PD	06-SEP-2000.									
XX	25-FEB-2000; 2000EP-0301439.									
XX	25-FEB-1999; 990S-0121825.									
PR	05-MAR-1999; 990S-0123180.									
PR	09-MAR-1999; 990S-0123548.									
PR	23-MAR-1999; 990S-0125788.									
PR	25-MAR-1999; 990S-0126264.									
PR	29-MAR-1999; 990S-0126785.									
PR	01-APR-1999; 990S-0127462.									
PR	06-APR-1999; 990S-0128234.									
PR	08-APR-1999; 990S-0128714.									
PR	16-APR-1999; 990S-0129845.									
PR	19-APR-1999; 990S-0130077.									
PR	21-APR-1999; 990S-0130449.									
PR	23-APR-1999; 990S-0130510.									
PR	23-APR-1999; 990S-0130891.									
PR	28-APR-1999; 990S-0131449.									
PR	30-APR-1999; 990S-0132048.									
PR	30-APR-1999; 990S-0132407.									
PR	04-MAY-1999; 990S-0132484.									
PR	05-MAY-1999; 990S-0132485.									
PR	06-MAY-1999; 990S-0132486.									
PR	06-MAY-1999; 990S-0132487.									
PR	07-MAY-1999; 990S-0132863.									
PR	11-MAY-1999; 990S-0134256.									
PR	14-MAY-1999; 990S-0134218.									
PR	14-MAY-1999; 990S-0134219.									

PR	01-JUN-1999;	9905-0137222;
PR	03-JUN-1999;	9905-0137528;
PR	04-JUN-1999;	9905-0137502;
PR	07-JUN-1999;	9905-0137724;
PR	08-JUN-1999;	9905-0138094;
PR	10-JUN-1999;	9905-0138504;
PR	10-JUN-1999;	9905-0138647;
PR	14-JUN-1999;	9905-0139119;
PR	14-JUN-1999;	9905-0139452;
PR	16-JUN-1999;	9905-0139453;
PR	16-JUN-1999;	9905-0139457;
PR	17-JUN-1999;	9905-0139459;
PR	18-JUN-1999;	9905-0139454;
PR	18-JUN-1999;	9905-0139455;
PR	18-JUN-1999;	9905-0139456;
PR	18-JUN-1999;	9905-0139457;
PR	18-JUN-1999;	9905-0139458;
PR	18-JUN-1999;	9905-0139459;
PR	18-JUN-1999;	9905-0139460;
PR	18-JUN-1999;	9905-0139461;
PR	18-JUN-1999;	9905-0139462;
PR	23-JUN-1999;	9905-0140353;
PR	23-JUN-1999;	9905-0140354;
PR	24-JUN-1999;	9905-0140695;
PR	28-JUN-1999;	9905-0140823;
PR	29-JUN-1999;	9905-0140991;
PR	30-JUN-1999;	9905-0141287;
PR	01-JUL-1999;	9905-0141842;
PR	01-JUL-1999;	9905-0142015;
PR	02-JUL-1999;	9905-0142054;
PR	06-JUL-1999;	9905-0142390;
PR	08-JUL-1999;	9905-0142803;
PR	09-JUL-1999;	9905-0142920;
PR	12-JUL-1999;	9905-0143297;
PR	13-JUL-1999;	9905-0143542;
PR	14-JUL-1999;	9905-0143624;
PR	15-JUL-1999;	9905-0144005;
PR	16-JUL-1999;	9905-0144085;
PR	16-JUL-1999;	9905-0144086;
PR	19-JUL-1999;	9905-0144335;
PR	19-JUL-1999;	9905-0144336;
PR	19-JUL-1999;	9905-0144337;
PR	19-JUL-1999;	9905-0144338;
PR	20-JUL-1999;	9905-0144632;
PR	20-JUL-1999;	9905-0144884;
PR	21-JUL-1999;	9905-0144814;
PR	21-JUL-1999;	9905-0145086;
PR	21-JUL-1999;	9905-0145087;
PR	22-JUL-1999;	9905-0145088;
PR	22-JUL-1999;	9905-0145087;
PR	22-JUL-1999;	9905-0145089;
PR	22-JUL-1999;	9905-0145109;
PR	23-JUL-1999;	9905-0145145;
PR	23-JUL-1999;	9905-0145145;
PR	23-JUL-1999;	9905-0145276;
PR	26-JUL-1999;	9905-0145274;
PR	27-JUL-1999;	9905-0145213;
PR	27-JUL-1999;	9905-0145216;
PR	27-JUL-1999;	9905-0145218;
PR	27-JUL-1999;	9905-0145219;
PR	28-JUL-1999;	9905-0145951;
PR	02-AUG-1999;	9905-0146386;
PR	02-AUG-1999;	9905-0146388;
PR	03-AUG-1999;	9905-0146389;
PR	03-AUG-1999;	9905-0147038;
PR	04-AUG-1999;	9905-0147204;
PR	04-AUG-1999;	9905-0147302;

PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142920.
PR 13-JUL-1999; 99US-0142977.
PR 14-JUL-1999; 99US-0143542.
PR 15-JUL-1999; 99US-0143624.
PR 16-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145921.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147204.

PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0148368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151338.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154479.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 26-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

PR	04-AUG-1999;	9905-0147204.	PR	29-OCT-1999;	9905-0162142.
PR	04-AUG-1999;	9905-0147302.	PR	05-MAR-1999;	9905-0121825.
PR	05-AUG-1999;	9905-0147192.	PR	09-MAR-1999;	9905-0123548.
PR	05-AUG-1999;	9905-0147260.	PR	23-MAR-1999;	9905-0125788.
PR	06-AUG-1999;	9905-0147303.	PR	25-MAR-1999;	9905-0126264.
PR	06-AUG-1999;	9905-0147416.	PR	29-MAR-1999;	9905-0126785.
PR	09-AUG-1999;	9905-0147493.	PR	01-APR-1999;	9905-0127462.
PR	09-AUG-1999;	9905-0147935.	PR	06-APR-1999;	9905-0128234.
PR	10-AUG-1999;	9905-0148171.	PR	08-APR-1999;	9905-0128714.
PR	11-AUG-1999;	9905-0148319.	PR	16-APR-1999;	9905-0129845.
PR	12-AUG-1999;	9905-0148341.	PR	19-APR-1999;	9905-0130077.
PR	13-AUG-1999;	9905-0148664.	PR	21-APR-1999;	9905-0130449.
PR	16-AUG-1999;	9905-0149368.	PR	23-APR-1999;	9905-0130510.
PR	17-AUG-1999;	9905-0149175.	PR	28-APR-1999;	9905-0130891.
PR	18-AUG-1999;	9905-0149426.	PR	30-APR-1999;	9905-0131449.
PR	20-AUG-1999;	9905-0149722.	PR	04-MAY-1999;	9905-0132407.
PR	20-AUG-1999;	9905-0149723.	PR	05-MAY-1999;	9905-0132484.
PR	20-AUG-1999;	9905-0149929.	PR	06-MAY-1999;	9905-0132485.
PR	23-AUG-1999;	9905-0149902.	PR	07-MAY-1999;	9905-0132487.
PR	23-AUG-1999;	9905-0149930.	PR	11-MAY-1999;	9905-0132863.
PR	25-AUG-1999;	9905-0150566.	PR	14-MAY-1999;	9905-0134256.
PR	26-AUG-1999;	9905-0150884.	PR	14-MAY-1999;	9905-0134219.
PR	27-AUG-1999;	9905-0151065.	PR	14-MAY-1999;	9905-0134221.
PR	27-AUG-1999;	9905-0151066.	PR	18-MAY-1999;	9905-0134370.
PR	27-AUG-1999;	9905-0151080.	PR	19-MAY-1999;	9905-0134768.
PR	30-AUG-1999;	9905-0151303.	PR	20-MAY-1999;	9905-0134941.
PR	31-AUG-1999;	9905-0151438.	PR	21-MAY-1999;	9905-0135124.
PR	01-SEP-1999;	9905-0151930.	PR	24-MAY-1999;	9905-0135353.
PR	07-SEP-1999;	9905-0152363.	PR	25-FEB-2000;	2000EP-0301439.
PR	10-SEP-1999;	9905-0153070.	XX	05-MAR-1999;	9905-0121825.
PR	13-SEP-1999;	9905-0153758.	XX	09-MAR-1999;	9905-0123548.
PR	15-SEP-1999;	9905-0154018.	XX	23-MAR-1999;	9905-0125788.
PR	16-SEP-1999;	9905-0154039.	XX	25-MAR-1999;	9905-0126264.
PR	20-SEP-1999;	9905-0154779.	XX	29-MAR-1999;	9905-0126785.
PR	22-SEP-1999;	9905-0155139.	XX	01-APR-1999;	9905-0127462.
PR	23-SEP-1999;	9905-0155486.	XX	06-APR-1999;	9905-0128234.
PR	24-SEP-1999;	9905-0155659.	XX	08-APR-1999;	9905-0128714.
PR	28-SEP-1999;	9905-0156458.	XX	16-APR-1999;	9905-0129845.
PR	29-SEP-1999;	9905-0156596.	XX	19-APR-1999;	9905-0130077.
PR	04-OCT-1999;	9905-0157117.	XX	21-APR-1999;	9905-0130449.
PR	05-OCT-1999;	9905-0157753.	XX	23-APR-1999;	9905-0130510.
PR	06-OCT-1999;	9905-0157865.	XX	28-APR-1999;	9905-0130891.
PR	07-OCT-1999;	9905-0158029.	XX	30-APR-1999;	9905-0131449.
PR	08-OCT-1999;	9905-0158232.	XX	04-MAY-1999;	9905-0132407.
PR	12-OCT-1999;	9905-0158369.	XX	05-MAY-1999;	9905-0132484.
PR	13-OCT-1999;	9905-0159283.	XX	06-MAY-1999;	9905-0132485.
PR	13-OCT-1999;	9905-0159294.	XX	07-MAY-1999;	9905-0132487.
PR	13-OCT-1999;	9905-0159295.	XX	11-MAY-1999;	9905-0132863.
PR	14-OCT-1999;	9905-0159329.	XX	14-MAY-1999;	9905-0134256.
PR	14-OCT-1999;	9905-0159330.	XX	14-MAY-1999;	9905-0134219.
PR	14-OCT-1999;	9905-0159331.	XX	14-MAY-1999;	9905-0134221.
PR	14-OCT-1999;	9905-0159637.	XX	18-MAY-1999;	9905-0134370.
PR	14-OCT-1999				

PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 28-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 15-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.

PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2002, 16:32:05 ; Search time 30.41 Seconds
(without alignments)
591.712 Million cell updates/sec

Title: US-09-676-718a-1

Perfect score: 162
Sequence: 1 MAAGPSGCLVPAGFKRLLLA.....ILKNWTSVEEELSEKLERI 162

Scoring table:
Gapop 60.0 , Gapext 60.0

Searched: 747574 seqs, 111073796 residues

Word size : 0

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

A.Genesec 032802:*
1: /SIDS1/gcgdata/hold-genesec/genesecp-emb1/AA1980.DAT:*
2: /SIDS1/gcgdata/hold-genesec/genesecp-emb1/AA1981.DAT:*
3: /SIDS1/gcgdata/hold-genesec/genesecp-emb1/AA1982.DAT:*
4: /SIDS1/gcgdata/hold-genesec/genesecp-emb1/AA1983.DAT:*
5: /SIDS1/gcgdata/hold-genesec/genesecp-emb1/AA1984.DAT:*
6: /SIDS1/gcgdata/hold-genesec/genesecp-emb1/AA1985.DAT:*
7: /SIDS1/gcgdata/hold-genesec/genesecp-emb1/AA1986.DAT:*
8: /SIDS1/gcgdata/hold-genesec/genesecp-emb1/AA1987.DAT:*
9: /SIDS1/gcgdata/hold-genesec/genesecp-emb1/AA1988.DAT:*
10: /SIDS1/gcgdata/hold-genesec/genesecp-emb1/AA1989.DAT:*
11: /SIDS1/gcgdata/hold-genesec/genesecp-emb1/AA1990.DAT:*
12: /SIDS1/gcgdata/hold-genesec/genesecp-emb1/AA1991.DAT:*
13: /SIDS1/gcgdata/hold-genesec/genesecp-emb1/AA1992.DAT:*
14: /SIDS1/gcgdata/hold-genesec/genesecp-emb1/AA1993.DAT:*
15: /SIDS1/gcgdata/hold-genesec/genesecp-emb1/AA1994.DAT:*
16: /SIDS1/gcgdata/hold-genesec/genesecp-emb1/AA1995.DAT:*
17: /SIDS1/gcgdata/hold-genesec/genesecp-emb1/AA1996.DAT:*
18: /SIDS1/gcgdata/hold-genesec/genesecp-emb1/AA1997.DAT:*
19: /SIDS1/gcgdata/hold-genesec/genesecp-emb1/AA1998.DAT:*
20: /SIDS1/gcgdata/hold-genesec/genesecp-emb1/AA1999.DAT:*
21: /SIDS1/gcgdata/hold-genesec/genesecp-emb1/AA2000.DAT:*
22: /SIDS1/gcgdata/hold-genesec/genesecp-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Query of match SUMMARIES

Result No.	Score	Match Length	DB ID	Description
1	161	99.4	162	21 AAY32112 Human 15 kDa selen
2	95	58.6	162	21 AAY32113 Mouse 15 kDa selen
3	77	47.5	92	20 AAM67876 Human secreted pro
4	77	47.5	95	21 AAY65395 Human 5' EST relat
5	77	47.5	124	21 AAB57132 Human prostate can
6	60	37.0	80	22 AAB75288 Gene 1 human secre
7	60	37.0	60	22 AAB75289 Human secreted pro
8	34	21.0	34	21 AAB44006 Human cancer assoc
9	28	17.3	50	21 AAY65389 Human 5' EST relat
10	8	4.9	142	22 ABB63319 Drosophila melanog
11	8	4.9	303	22 AAU46002 Propionibacterium

12	7	4.3	46	20 AAY27650
13	7	4.3	120	21 AAU49944
14	7	4.3	149	21 AAG08634
15	7	4.3	151	12 AAR13304
16	7	4.3	155	21 AAG08633
17	7	4.3	155	21 AAG43258
18	7	4.3	156	22 AAM42036
19	7	4.3	158	21 AAG08632
20	7	4.3	159	21 AAG43257
21	7	4.3	159	21 AAG43256
22	7	4.3	190	22 AAG07527
23	7	4.3	281	21 AAG07226
24	7	4.3	319	21 AAG07225
25	7	4.3	346	18 AAM23088
26	7	4.3	364	14 AAR31014
27	7	4.3	424	22 AAM51207
28	7	4.3	426	22 AAB67773
29	7	4.3	454	22 AAB59876
30	7	4.3	492	12 AAR13336
31	7	4.3	503	19 AAM53332
32	7	4.3	510	19 AAM38289
33	7	4.3	541	12 AAR13296
34	7	4.3	660	18 AAM44120
35	7	4.3	734	22 AAG91651
36	7	4.3	734	22 AAB76657
37	7	4.3	809	22 AAG09902
38	7	4.3	836	16 AAR79835
39	7	4.3	1091	12 AAR13303
40	7	4.3	1190	12 AAR13308
41	7	4.3	1255	22 AAG84961
42	7	4.3	1284	22 AAB62713
43	7	4.3	1476	20 AAY33968
44	7	4.3	1479	12 AAR11602
45	7	4.3	1479	12 AAR13231

ALIGNMENTS

RESULT 1	
AAV32112	
ID	AAV32112 standard; Protein: 162 AA.
XX	
AC	AAV32112;
XX	
DT	01-FEB-2000 (first entry)
XX	
DE	Human 15 kDa selenoprotein.
XX	
KW	Selenoprotein; selenum; differential expression; tumour; human;
KW	prostate cancer; diagnosis; polymorphism.
XX	
OS	Homo sapiens.
XX	
PH	
FT	Key
FT	Peptide
FT	Location/Qualifiers
FT	1..26
FT	/note="signal peptide"
FT	27..162
FT	/note="mature protein (specifically claimed in
FT	Claim 12")
FT	Modified-site
FT	93
FT	/label="OTHER
FT	/note="selenocysteine"
PN	W09951637-AL.
XX	
PD	14-OCT-1999.
XX	
PF	06-APR-1999;
XX	99WO-US07560.
XX	
PR	06-APR-1998;
XX	98US-0080850.
XX	
PA	(USSH) US DEPT HEALTH & HUMAN SERVICES.

Human secreted pro
Propionibacterium
Arabidopsis thalia
CFRR 556 del A. H
Arabidopsis thalia
Arabidopsis thalia
Human polypeptide
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Novel human diagno
Arabidopsis thalia
Pyrodicticum sp. es
Prod. of a probe o
Rat Csk binding pr
Drosophila melanog
CFRR Q493X. Homo
Cryptosporidium pa
Murine N-acetylgl
CFRR G542X. Homo
Human ADAM 12 pred
C glutamicum prote
Cornebacterium g1
Novel human diagno
Truncated cystic f
CFRR Y1092X. Homo
CFRR 3659 del C.
Shrimp white spot
Drosophila melanog
CFRR protein sequ
Mutant cystic fibr
CFRR delta 1507.

XX Gladyshev VN, Wootton JC, Hatfield DL, Jeang K;
PI WPI: 2000-013045/01.
DR N-PSDB: AA234464.
XX
PT New isolated selenoprotein polypeptides, used to develop products for
PT detecting susceptibility to or treating cancers e.g. prostate cancers
PT
PS Claim 1; Page 55; 67pp: English.
XX
CC The present sequence represents a human 15 kDa selenoprotein that
CC is differentially expressed in cancer cells, such as cancer cells.
CC It includes a selenocysteine residue at position 93 that is encoded
CC by a TGA codon in the cDNA (see AA234464). Polymorphisms in the
CC 15 kDa selenoprotein gene are associated with susceptibility to
CC cancer. Claimed methods of detecting the presence, or of
CC quantifying the level of expression, of the 15 kDa selenoprotein
CC involve contacting a sample with a specific binding agent, such as
CC a polyclonal antibody or monoclonal antibody, that specifically
CC binds to the 15 kDa selenoprotein. A claimed method for dietary
CC regulation comprises detecting an abnormally low expression of a 15
CC kDa selenoprotein in a subject and, if the level is below normal,
CC enhancing the level by providing additional selenium in the diet.
CC The susceptibility to cancer of a subject having an increased
CC predetermined genetic susceptibility to cancer is reduced by
CC administering the 15 kDa selenoprotein or by overexpressing the
CC selenoprotein using gene therapy.
XX
SQ Sequence 162 AA;
SO
Query Match 99.4%; Score 161; DB 21; Length 162;
Best Local Similarity 100.0%; Pred. No. 2.9e-150;
Matches 162; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAAPSGCIVAFKRIILAVIAVSAFSGAFSEACRELGFSNLLCSCDILGQNL 60
DB 1 MAAPSGCIVAFKRIILAVIAVSAFSGAFSEACRELGFSNLLCSCDILGQNL 60
QY 61 LQLPDPCGCCQEAQFETKRLVAGAILVCGXKLRPPVOVQAFVRSKPKLFRGLQIKY 120
DB 61 LQLPDPCGCCQEAQFETKRLVAGAILVCGXKLRPPVOVQAFVRSKPKLFRGLQIKY 120
QY 121 VRGSDPYVKLLDNGNIAEELSLKWMNTDVEEFLSEKLERI 162
DB 121 VRGSDPYVKLLDNGNIAEELSLKWMNTDVEEFLSEKLERI 162
RESULT 2
ID AAY32113 standard; Protein: 162 AA.
XX
AC AAY32113;
XX
DT 01-FEB-2000 (first entry)
XX
DE Mouse 15 kDa selenoprotein.
XX
KM Selenoprotein; selenium; differential expression; tumour; mouse;
KM prostate cancer; diagnosis; polymorphism.
XX
OS Mus musculus.
XX
XX Key Location/Qualifiers
FT Modified-site 93
FT /label=OTHER
FT /note="selenocysteine"
XX
PN MO9951637-A1.
XX
PD 14-OCT-1999.

XX
PF 06-APR-1999; 99WO-US07560.
XX
PR 06-APR-1998; 98US-0080850.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Gladyshev VN, Wootton JC, Hatfield DL, Jeang K;
XX
DR WPI: 2000-013045/01.
DR N-PSDB: AA234469.
XX
PT New isolated selenoprotein polypeptides, used to develop products for
PT detecting susceptibility to or treating cancers e.g. prostate cancers
PT
PS Claim 1; Page 60; 67pp: English.
XX
CC The present sequence represents a mouse 15 kDa selenoprotein that
CC is differentially expressed in cancer cells, such as cancer cells.
CC It includes a selenocysteine residue at position 93 that is encoded
CC by a TGA codon in the cDNA (see AA234468). Polymorphisms in the
CC human 15 kDa selenoprotein gene are associated with susceptibility
CC to cancer. Claimed methods of detecting the presence, or of
CC quantifying the level of expression, of the 15 kDa selenoprotein
CC involve contacting a sample with a specific binding agent, such as
CC a polyclonal antibody or monoclonal antibody, that specifically
CC binds to the 15 kDa selenoprotein. A claimed method for dietary
CC regulation comprises detecting an abnormally low expression of a
CC mammalian 15 kDa selenoprotein in a subject and, if the level is
CC below normal, enhancing the level by providing additional selenium
CC in the diet. The susceptibility to cancer of a subject having an
CC increased predetermined genetic susceptibility to cancer is reduced
CC by administering a mammalian 15 kDa selenoprotein or by
CC overexpressing the selenoprotein using gene therapy.
XX
SQ Sequence 162 AA;
SO
Query Match 58.6%; Score 95; DB 21; Length 162;
Best Local Similarity 100.0%; Pred. No. 2.4e-85;
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 67 CRCCQEAQFETKRLVAGAILVCGXKLRPPVOVQAFVRSKPKLFRGLQIKYVRS 126
DB 67 CRCCQEAQFETKRLVAGAILVCGXKLRPPVOVQAFVRSKPKLFRGLQIKYVRS 126
QY 127 VKLLDNGNIAEELSLKWMNTDVEEFLSEKLERI 162
DB 127 VKLLDNGNIAEELSLKWMNTDVEEFLSEKLERI 162
RESULT 3
ID AAW67876 standard; Protein: 92 AA.
XX
AC AAW67876;
XX
DT 25-MAR-1999 (first entry)
XX
DE Human secreted protein encoded by gene 70 clone HMSG42.
XX
KM Human; secreted protein; fusion protein; gene therapy; protein therapy;
KM diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KM developmental abnormality; foetal deficiency; blood; allergy; renal;
KM immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KM inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KM cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KM osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KM endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX
OS Homo sapiens.
XX

PN W09842738-A1.
 XX 01-OCT-1998.
 PD
 XX
 PF 19-MAR-1998; 98WO-US05311.
 XX
 PR 30-MAY-1997; 97US-0050937.
 PR 21-MAR-1997; 97US-0041276.
 PR 21-MAR-1997; 97US-0041277.
 PR 21-MAR-1997; 97US-0041281.
 PR 21-MAR-1997; 97US-0042344.
 PR 30-MAY-1997; 97US-0048069.
 PR 30-MAY-1997; 97US-0048094.
 PR 30-MAY-1997; 97US-0048095.
 PR 30-MAY-1997; 97US-0048096.
 PR 30-MAY-1997; 97US-0048099.
 PR 30-MAY-1997; 97US-0048131.
 PR 30-MAY-1997; 97US-0048135.
 PR 30-MAY-1997; 97US-0048154.
 PR 30-MAY-1997; 97US-0048160.
 PR 30-MAY-1997; 97US-0048186.
 PR 30-MAY-1997; 97US-0048187.
 PR 30-MAY-1997; 97US-0048188.
 PR 30-MAY-1997; 97US-0048350.
 PR 30-MAY-1997; 97US-0048351.
 PR 30-MAY-1997; 97US-0048352.
 PR 30-MAY-1997; 97US-0048355.
 PR 05-AUG-1997; 97US-0054804.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Brewer LA, Duan R, Ebner R, Ferrle AM, Florence KA;
 PI Greene JM, Hu JS, Lafleur DM, Moore PA, Ni J, Olsen HS;
 PI Rosen CA, Ruben SM, Shi Y, Young P;
 XX
 DR MPI: 1999-070066/06.
 DR N-PSDB; AAX00680.
 XX
 PT New isolated human genes and the secreted polypeptides they encode -
 PT useful for diagnosis and treatment of e.g. cancers, neurological
 PT disorders, immune diseases, inflammation or blood disorders
 XX
 PS Claim 11: Page 310; 385pp; English.
 XX
 CC This sequence represents a secreted human protein encoded by the gene
 CC clone detailed in the descriptor line. The gene can be used to generate
 CC fusion proteins by linking to the gene to a human immunoglobulin Fc
 CC portion (e.g. AAX00602) for increasing the stability of the fused
 CC protein as compared to the human protein only.
 CC The invention relates to 87 novel genes and their fragments (nucleic
 CC acid sequences: AAX00611-X00724; amino acid sequences AAW67807-M68004)
 CC which are useful for preventing, treating or ameliorating medical
 CC conditions e.g. by protein or gene therapy. Also, pathological
 CC conditions can be diagnosed by determining the amount of the new
 CC polypeptides in a sample or by determining the presence of mutations in
 CC the new polynucleotides. Specific uses are described for each of the 87
 CC polynucleotides, based on which tissues they are most highly expressed in
 CC (see AAX00611 for described uses).
 CC
 XX
 SQ Sequence 92 AA;

Query Match 47.5%; Score 77; DB 20; Length 92;
 Best Local Similarity 100.0%; Pred. No. 7.2e-68;
 Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 16 RLILATVLAQAVSAFGAFESSEACRELGFSNNLCCSDLLGQNLQLQDPDRCGCCQEEA 75
 DB 16 RLILATVLAQAVSAFGAFESSEACRELGFSNNLCCSDLLGQNLQLQDPDRCGCCQEEA 75
 OY 76 QFETKRLVAGAILVEVC 92
 DB 76 QFETKRLVAGAILVEVC 92

RESULT 4
 ID AAY65395
 XX AAY65395 standard; Protein; 95 AA.
 AC AAY65395;
 XX
 DT 01-FEB-2000 (first entry)
 XX
 DE Human 5' EST related polypeptide SEQ ID NO:1556.
 XX
 KW Human; 5' EST; expressed sequence tag; secreted protein; diagnosis;
 KW gene therapy; chromosome mapping; upstream regulatory sequence;
 KW forensic; location; development; protein synthesis; stability;
 KW regulation; identification.
 XX
 OS Homo sapiens.
 XX
 PN W09953051-A2.
 XX
 PD 21-OCT-1999.
 XX
 PF 09-APR-1999; 99WO-IB00712.
 XX
 PR 09-APR-1998; 98US-0057719.
 PR 28-APR-1998; 98US-0069047.
 XX
 PA (GEST) GENSET.
 XX
 PI Dumas Milne Edwards J, Duclert A, Giordano J;
 PI MPI: 2000-038446/03.
 DR N-PSDB; AAZ43009.
 XX
 PT Novel secreted protein 5' expressed sequence tag sequences used in
 PT diagnostic, forensic, gene therapy, and chromosome mapping procedures
 XX
 PS Claim 3: Page 818; 837pp; English.
 XX
 CC AAZ42265 to AAZ43075 represent novel 5' expressed sequence tag (EST)
 CC sequences, corresponding to human secreted proteins. AAY64651 to
 CC AAY65438 represent the EST-related proteins corresponding to AAZ42265 to
 CC AAZ43052. The 5' ESTs can be used for producing secreted human gene
 CC products. They can be used to identify and isolate 5' untranslated
 CC regions (UTRs) and upstream regulatory regions which control the
 CC location, development stage, rate, and quantity of protein synthesis, as
 CC well as stability of mRNA. The ESTs are also useful as probes for
 CC chromosome mapping, and to obtain full length cDNA clones. The ESTs can
 CC also be used in forensic procedures to identify individuals, or in
 CC diagnostic procedures to identify individuals having genetic diseases
 CC resulting from abnormal gene expression. The products may also be used in
 CC gene therapy protocols. The nucleic acids encoding signal peptides can be
 CC used for directing extracellular secretion of a polypeptide or the
 CC insertion of a polypeptide into a membrane, or importing a polypeptide
 CC into a cell. The proteins encoded by the EST sequences may be useful in
 CC treating a variety of human conditions. Secreted proteins have
 CC therapeutic value, and the identification of new secreted proteins is
 CC valuable. AAZ42249 to AAZ42264 and AAY64644 to AAY64650 represent
 CC sequences used in the exemplification of the present invention.
 CC
 XX
 SQ Sequence 95 AA;

Query Match 47.5%; Score 77; DB 21; Length 95;
 Best Local Similarity 100.0%; Pred. No. 7.4e-68;
 Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 16 RLILATVLAQAVSAFGAFESSEACRELGFSNNLCCSDLLGQNLQLQDPDRCGCCQEEA 75
 DB 19 RLILATVLAQAVSAFGAFESSEACRELGFSNNLCCSDLLGQNLQLQDPDRCGCCQEEA 78
 OY 76 QFETKRLVAGAILVEVC 92

Db 79 qfckklyagailevcg 95
|||||
RESULT 5
ID AAB57132 standard; Protein: 124 AA.
XX AAB57132;
XX
DT 13-MAR-2001 (first entry)
XX
DE Human prostate cancer antigen protein sequence SEQ ID NO:1710.
XX
KW Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
KW neuroprotective; cytosolic; cardioprotective; immunomodulatory; muscular;
KW vulnary; gastrointestinal; nephrotoxic; antineoplastic; gynaecological;
KW antibacterial; gene therapy; neural; immune; reproductive; renal;
KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
KW wound; infectious disease.
XX
OS Homo sapiens.
XX
PN WO20005174-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05988.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI: 2000-587513/55.
DR N-PSDB: AAF16335.
XX
PT Prostate cancer associated gene sequences, referred to as prostate
PT cancer antigens, useful for treatment, prevention, and diagnosis of
PT disorders such as prostate cancer -
XX
XX
PS Claim 11; Page 2183-2184; 2338pp: English.
XX
CC AAF15566 to AAF16505 encode the human prostate cancer associated
CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
CC The prostate cancer antigens can have neuroprotective, cytosolic,
CC cardioprotective, immunomodulatory, muscular, vulnary, gastrointestinal,
CC nephrotoxic, antineoplastic, gynaecological and antibacterial activities,
CC and can be used in gene therapy. The prostate cancer antigen
CC polynucleotides may be used for detection of prostate cancer, chromosome
CC identification, as chromosome markers, and for numerous other diagnostic
CC or research purposes. The prostate cancer antigens may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders, wounds, and infectious diseases. AAF15506 to AAF16514 to
CC AAB57303 represent sequences used in the exemplification of the present
CC invention.
XX
SQ Sequence 124 AA;
Query Match 47.5%; Score 77; DB 21; Length 124;
Best Local Similarity 100.0%; Pred. No. 9.3e-68;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 RLILATYQVAVSARFAGFSSACRELGSSMLSSCDLLQFNLQLDPCCRGCCQGEA 75
Db 48 RLILATYQVAVSARFAGFSSACRELGSSMLSSCDLLQFNLQLDPCCRGCCQGEA 107
QY 76 QEFTRKLYAGAILLEVC 92
|||||

Db 108 qfckklyagailevcg 124
|||||
RESULT 6
ID AAB75288 standard; Protein: 60 AA.
XX AAB75288;
XX
DT 03-APR-2001 (first entry)
XX
DE Gene 1 human secreted protein homologous amino acid sequence #107.
XX
KW Human; immunosuppressive; antiarthritic; antirheumatic; nootropic;
KW antiproliferative; cytosolic; cardiatic; vasotropic; cerebroprotective;
KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
KW vulnary; autoimmune disease; hyperproliferative disorder; cancer;
KW cardiovascular disorder; cerebrovascular disorder; infection;
KW nervous system disorder; ocular disorder; chemotaxis; food additive;
KW secreted protein.
XX
OS Homo sapiens.
XX
PN WO200077021-A1.
XX
PD 21-DEC-2000.
XX
PF 01-JUN-2000; 2000WO-US15135.
XX
PR 11-JUN-1999; 99US-0138632.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
XX
PI Rosen CA, Ruben SM, Komatsoullis GA;
XX
DR WPI: 2001-071257/08.
XX
PT Nucleic acid molecules encoding human secreted proteins, used in
PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
PT Parkinson's diseases and cancers -
XX
XX
PS Disclosure; Page 10; 530pp: English.
XX
CC This invention relates to polynucleotide sequences AAF63789 - AAF63836
CC which encode human secreted proteins AAB75260 - AAB75287. Included in the
CC invention are protein sequences AAB75288 - AAB75341 which are fragments
CC of the secreted proteins and amino acid sequences with which these
CC fragments share homology. Examples of the activities of the proteins and
CC polynucleotides and the activities of their agonists and antagonists
CC include, immunosuppressive; antiarthritic; antirheumatic;
CC antiproliferative; cytosolic; cardiatic; vasotropic; cerebroprotective;
CC nootropic; neuroprotective; antibacterial; virucide; fungicide;
CC ophthalmological; and vulnary activity. The protein and polynucleotide
CC sequences, their agonists and antagonists may be useful for treating,
CC preventing and diagnosing diseases and disorders such as autoimmune
CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders
CC e.g. neoplasms of the breast or liver, cardiovascular disorders
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischemia,
CC angiogenesis, nervous system disorders e.g. Alzheimer's disease,
CC infections caused by bacteria, viruses and fungi and ocular disorders
CC e.g. corneal infection. The polypeptides can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities. Included in the invention are
CC oligonucleotides AAF63780 - AAF63788 and peptide AAB75239 which are used
CC in the identification and characterisation of the DNA and protein
CC sequences of the invention.
XX
SQ Sequence 60 AA;

Query Match 37.0%; Score 60; DB 22; Length 60;
 Best Local Similarity 100.0%; Pred. No. 2.6e-51;
 Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 103 AFVRSDBPKLFRGLQIKYVSGSDPVKLLDDNGNIAEELSLKMNWDSEEFLEKLERI 162
 Db 1 afvrsdkpklfrglqikyvsgsdpvkllddngniaeelslkwnldsvseeflsekeri 60

RESULT 7
 AAB75289
 ID AAB75289 standard; Protein; 60 AA.
 XX AAB75289;
 AC AAB75289;
 CC AAB75289;
 DD AAB75289;
 EE AAB75289;
 FF AAB75289;
 GG AAB75289;
 HH AAB75289;
 II AAB75289;
 JJ AAB75289;
 KK AAB75289;
 LL AAB75289;
 MM AAB75289;
 NN AAB75289;
 OO AAB75289;
 PP AAB75289;
 QQ AAB75289;
 RR AAB75289;
 SS AAB75289;
 TT AAB75289;
 UU AAB75289;
 VV AAB75289;
 WW AAB75289;
 XX AAB75289;
 YY AAB75289;
 ZZ AAB75289;

Human secreted protein sequence encoded by gene 1 SEQ ID NO:108.

Human; immunosuppressive; antiarthritic; antirheumatic; neutrotropic; antiproliferative; cytostatic; cardiant; vasotrophic; cerebroprotective; neuroprotective; antibacterial; virucide; fungicide; ophthalmological; vulnary; autoimmune disease; hyperproliferative disorder; cancer; cardiovascular disorder; cerebrovascular disorder; infection; nervous system disorder; ocular disorder; chemotaxis; food additive; secreted protein.

Homo sapiens.

WO200077021-A1.

21-DEC-2000.

01-JUN-2000; 2000WO-US15135.

11-JUN-1999; 99US-0138632.

(HUMA-) HUMAN GENOME SCI INC.
 (ROSE/) ROSEN C A.

Rosen CA, Ruben SM, Komatsoulis GA;
 WPI: 2001-071257/08.
 N-PSDB: AAF63789.

Nucleic acid molecules encoding human secreted proteins, used in preventing, treating or ameliorating a disorder, e.g. Alzheimer's and Parkinson's diseases and cancers -

Disclosure: Page 10; 530pp; English.

This invention relates to polynucleotide sequences AAF63789 - AAF63836 which encode human secreted proteins AAB75260 - AAB75287. Included in the invention are protein sequences AAB75288 - AAB75341 which are fragments of the secreted proteins and amino acid sequences with which these fragments share homology. Examples of the activities of the proteins and polynucleotides and the activities of their agonists and antagonists include, immunosuppressive; antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant; vasotrophic; cerebroprotective; neuroprotective; antibacterial; virucide; fungicide; ophthalmological; and vulnary activity. The protein and polynucleotide sequences, their agonists and antagonists may be useful for treating, preventing and diagnosing diseases and disorders such as autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g. Alzheimer's disease, infections caused by bacteria, viruses and fungi and ocular disorders e.g. corneal infection. The polypeptides can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to suburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues and in chemotaxis. The

polypeptides can also be used as a food additive or preservative to increase or decrease storage capabilities. Included in the invention are oligonucleotides AAF63780 - AAF63788 and peptide AAB75239 which are used in the identification and characterisation of the DNA and protein sequences of the invention.

Sequence 60 AA;

Query Match 37.0%; Score 60; DB 22; Length 60;
 Best Local Similarity 100.0%; Pred. No. 2.6e-51;
 Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 103 AFVRSDBPKLFRGLQIKYVSGSDPVKLLDDNGNIAEELSLKMNWDSEEFLEKLERI 162
 Db 1 afvrsdkpklfrglqikyvsgsdpvkllddngniaeelslkwnldsvseeflsekeri 60

RESULT 8
 AAB44006
 ID AAB44006 standard; Protein; 34 AA.
 XX AAB44006;
 AC AAB44006;
 CC AAB44006;
 DD AAB44006;
 EE AAB44006;
 FF AAB44006;
 GG AAB44006;
 HH AAB44006;
 II AAB44006;
 JJ AAB44006;
 KK AAB44006;
 LL AAB44006;
 MM AAB44006;
 NN AAB44006;
 OO AAB44006;
 PP AAB44006;
 QQ AAB44006;
 RR AAB44006;
 SS AAB44006;
 TT AAB44006;
 UU AAB44006;
 VV AAB44006;
 WW AAB44006;
 XX AAB44006;
 YY AAB44006;
 ZZ AAB44006;

Human cancer associated protein sequence SEQ ID NO:1451.

Human; cancer associated gene; cancer antigen; detection; cancer; diagnosis; cytostatic; proliferative; vulnary; immunomodulator; antidiabetic; antiaesthetic; antirheumatic; antiarthritic; antiviral; antinflammatory; antithyroid; antiallergic; antibacterial; cardiant; dermatological; neuroprotective; thrombolytic; coagulant; neutrotropic; vasotrophic; antiproliferative; antidiabetic; gene therapy; inflammation; immune disorder; haematopoietic cell disorder; autoimmune disorder; allergic reaction; graft versus host disease; organ rejection; haemostatic; thrombolytic; cardiovascular disorder; infection; neurological disease; drug screening.

Homo sapiens.

WO200055350-A1.

21-SEP-2000.

08-MAR-2000; 2000WO-US05882.

12-MAR-1999; 99US-0124270.

(HUMA-) HUMAN GENOME SCI INC.
 (ROSEN/) ROSEN C A.

Rosen CA, Ruben SM;
 WPI: 2000-587533/55.
 N-PSDB: AAC78215.

Novel isolated nucleic acids comprising sequences encoding peptides useful for treating or diagnosing e.g. cancer -

Claim 11; Page 2132; 2352pp; English.

AAC77607 to AAC78448 encode the human cancer associated proteins given in AAB43398 to AAB44239. The proteins can have activities based on the tissues and cells the genes are expressed in. Example of activities include: cytostatic; proliferative; vulnary; immunomodulator; antidiabetic; antiaesthetic; antirheumatic; antiarthritic; antinflammatory; antithyroid; antiallergic; antibacterial; antiviral; dermatological; neuroprotective; cardiant; thrombolytic; coagulant; neutrotropic; vasotrophic; antiproliferative; antidiabetic; gene therapy; inflammation; immune disorder; haematopoietic cell disorder; autoimmune disorder; allergic reaction; graft versus host disease; organ rejection; haemostatic; thrombolytic; cardiovascular disorder; infection; neurological disease; drug screening.

Polynucleotides, polypeptides, antibodies, agonists and antagonists from the present invention may be used to treat immune disorders by activating

or inhibiting the proliferation, differentiation or mobilisation of immune cells, to treat disorders of haematopoietic cells, autoimmune disorders, allergic reactions, graft versus host disease and organ rejection, modulate haemostatic or thrombolytic activity, modulate inflammation, cancers, cardiovascular disorders, neurological disease and bacterial or viral infections. The peptides, nucleotides, antibodies, agonists and antagonists may be also be used in drug screens. AAC78449 to AAC78457 and AAB44240 represent sequences used in the exemplification of the present invention.

Sequence 34 AA:

Query Match 21.0%; Score 34; DB 21; Length 34;
Best Local Similarity 100.0%; Pred. No. 5.7e-26;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 KLIDNGNIAEELSLKWNTPVEEPLSEKLERI 162
|
Db 1 KLIDNGNIAEELSLKWNTPVEEPLSEKLERI 34

RESULT 9

AAV65389 standard; Protein; 50 AA.

AAV65389;

01-FEB-2000 (first entry)

Human 5' EST related polypeptide SEQ ID NO:1550.

Human; 5' EST; expressed sequence tag; secreted protein; diagnosis; gene therapy; chromosome mapping; upstream regulatory sequence; forensic; location; development; protein synthesis; stability; regulation; identification.

Homo sapiens.

WO9535051-A2.

21-OCT-1999.

09-APR-1999; 99WO-IB00712.

09-APR-1998; 98US-0057719.

28-APR-1998; 98US-0069047.

(GEST) GENSET.

Dumas Mline Edwards J, Duclert A, Giordano J;

WPI; 2000-038446/03.

N-PSDB; AAZ43003.

Novel secreted protein 5' expressed sequence tag sequences used in diagnostic, forensic, gene therapy, and chromosome mapping procedures

Claim 3; Page 816-817; 837pp; English.

AAZ42265 to AAZ43075 represent novel 5' expressed sequence tag (EST) sequences, corresponding to human secreted proteins. AAV64651 to AAV6538 represent the EST-related proteins corresponding to AAZ42265 to AAZ43052. The 5' ESTs can be used for producing secreted human gene products. They can be used to identify and isolate 5' untranslated regions (UTRs) and upstream regulatory regions which control the location, development stage, rate, and quantity of protein synthesis, as well as stability of mRNA. The ESTs are also useful as probes for chromosome mapping, and to obtain full length cDNA clones. The ESTs can also be used in forensic procedures to identify individuals, or in diagnostic procedures to identify individuals having genetic diseases resulting from abnormal gene expression. The products may also be used in gene therapy protocols. The nucleic acids encoding signal peptides can be

used for directing extracellular secretion of a polypeptide or the insertion of a polypeptide into a membrane, or importing a polypeptide into a cell. The proteins encoded by the EST sequences may be useful in treating a variety of human conditions. Secreted proteins have therapeutic value, and the identification of new secreted proteins is valuable. AAZ42249 to AAZ42264 and AAV6464 to AAV64650 represent sequences used in the exemplification of the present invention.

Sequence 50 AA:

Query Match 17.3%; Score 28; DB 21; Length 50;
Best Local Similarity 100.0%; Pred. No. 6.4e-20;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 KYRGSDPVLKLIDNGNIAEELSLKW 146
|
Db 20 KYRGSDPVLKLIDNGNIAEELSLKW 47

RESULT 10

ABB63319 standard; Protein; 142 AA.

ABB63319;

26-MAR-2002 (first entry)

Drosophila melanogaster polypeptide SEQ ID NO 16749.

Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.

Drosophila melanogaster.

WO200171042-A2.

27-SEP-2001.

23-MAR-2001; 2001WO-US09231.

23-MAR-2000; 2000US-191637P.

11-JUL-2000; 2000US-0614150.

(PEKE) PE CORP NY.

Venter JC, Adams M, Li PMD, Myers EW;

WPI; 2001-656860/75.

N-PSDB; ABL07422.

New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -

Disclosure; SEQ ID NO 16749; 21pp + Sequence Listing; English.

The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-ABBS72072).

The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 142 AA:

Query Match 4.9%; Score 8; DB 22; Length 142;

Best Local Similarity 100.0%; Pred. No. 7.6;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 116 IQIKYRG 123
 |||||
 Db 69 lqkyvrg 76

RESULT 11

AAU46002
 ID AAU46002 standard; Protein; 303 AA.

AC AAU46002;

DT 27-FEB-2002 (first entry)

DE Propionibacterium acnes immunogenic protein #6898.

XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
 KM dermatological; osteopathic; neuroprotectant.

XX Propionibacterium acnes.

PN WO200181581-A2.

PD 01-NOV-2001.

PF 20-APR-2001; 2001WO-US12865.

PR 21-APR-2000; 2000US-199047P.

PR 02-JUN-2000; 2000US-208841P.

PR 07-JUL-2000; 2000US-216747P.

XX (CORI-) CORIXA CORP.

XX Skelley YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;
 PI L'maisonneuve J, Zhang Y, Jen S, Carter D;

DR MPI: 2001-616774/71.

XX N-PSDB; AAS59529.

XX Propionibacterium acnes polypeptides and nucleic acids useful for
 PT vaccinating against and diagnosing infections, especially useful for
 PT treating acne vulgaris -

PS Example 1; SEQ ID NO 7197; 1066pp; English.

XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
 CC polypeptides. The proteins and their associated DNA sequences are used in
 CC the treatment, prevention and diagnosis of medical conditions caused by
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
 CC P. acnes is also involved in infections of bone, joints and the central
 CC nervous system, however it is particularly involved in the inflammatory
 CC lesions associated with acne vulgaris. A method for detecting the
 CC presence or absence of P. acnes in a patient comprises contacting a
 CC sample with a binding agent that binds to the proteins of the invention
 CC and determining the amount of bound protein in the sample. The
 CC polypeptides may be used as antigens in the production of antibodies
 CC specific for P. acnes proteins. These antibodies can be used to
 CC downregulate expression and activity of P. acnes polypeptides and
 CC therefore treat P. acnes infections. The antibodies may also be used as
 CC diagnostic agents for determining P. acnes presence, for example, by
 CC enzyme linked immunosorbent assay (ELISA).
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 303 AA;

Query Match 4.9%; Score 8; DB 22; Length 303;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 155 LSEKLERI 162
 |||||
 Db 167 lseklert 174

RESULT 12

AAZ27650
 ID AAZ27650 standard; Protein; 46 AA.

AC AAZ27650;

DT 30-JUL-1999 (first entry)

DE Human secreted protein encoded by gene No. 84.

XX Human; secreted protein; fusion protein; gene therapy; protein therapy;
 KW diagnosis; tissue; cancer; tumor; neurodegenerative disorder; leukemia;
 KW developmental abnormality; foetal deficiency; blood; allergy; renal;
 KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
 KW inflammation; ischemic shock; Alzheimer's disease; restenosis; AIDS;
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.

XX Homo sapiens.

PN WO9924836-A1.

PD 20-MAY-1999.

PF 04-NOV-1998; 98WO-US23435.

PR 17-NOV-1997; 97US-0066100.

PR 07-NOV-1997; 97US-0064900.

PR 07-NOV-1997; 97US-0064908.

PR 07-NOV-1997; 97US-0064911.

PR 07-NOV-1997; 97US-0064912.

PR 07-NOV-1997; 97US-0064983.

PR 07-NOV-1997; 97US-0064984.

PR 07-NOV-1997; 97US-0064985.

PR 07-NOV-1997; 97US-0064987.

PR 07-NOV-1997; 97US-0064988.

PR 17-NOV-1997; 97US-0066090.

PR 17-NOV-1997; 97US-0066094.

PR 17-NOV-1997; 97US-0066095.

PR 17-NOV-1997; 97US-0066089.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Carter KC, Edner R, Endress GA, Feng P, Janat F;
 PI Kyaw H, Lafleur DW, Moore PA, Ni J, Olsen HS, Rosen CA;
 PI Ruben SM, Shi Y, Soppet DR, Wei Y;

DR MPI: 1999-337740/28.

XX N-PSDB; AAX85016.

XX New human secreted proteins and coding sequences useful for treating
 PT disorders of the immune system and hyperproliferative disorders
 XX Claim 11; Page 397; 507pp; English.

XX This sequence represents a secreted human protein encoded by the gene
 CC clone detailed in the descriptor line. The gene can be used to generate
 CC fusion proteins by linking to the gene to a human immunoglobulin Fc
 CC portion (e.g. AAX84924) for increasing the stability of the fused
 CC protein as compared to the human protein only.
 CC The invention relates to 125 novel genes and their fragments (nucleic
 CC acid sequences: AAX84933-X85057; amino acid sequences AAZ7567-727933)
 CC which are useful for preventing, treating or ameliorating medical

CC conditions e.g. by protein or gene therapy. Also, pathological
CC conditions can be diagnosed by determining the amount of the new
CC polypeptides in a sample or by determining the presence of mutations in
CC the new polynucleotides. Specific uses are described for each of the 125
CC polynucleotides, based on which tissues they are most highly expressed in
CC (see AAX04933 for described uses).

CC
XX
SQ Sequence 46 AA;

Query Match 4.3%; Score 7; DB 20; Length 46;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 122 RGSDEVL 128
|111111|
Db 18 rgsdpl 24

RESULT 13
AAU49944 standard; Protein; 120 AA.
XX
AC AAU49944;
XX
DT 27-FEB-2002 (first entry)
XX
DE Propionibacterium acnes immunogenic protein #10840.
XX
XX SAPHO syndrome; synovitis; acne; pustulosis; hyperostosis; osteomyelitis;
KM uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KM inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KM dermatological; osteopathic; neuroprotectant.
XX
XX Propionibacterium acnes.
XX
XX W0200181581-A2.
XX
PD 01-NOV-2001.
XX
PF 20-APR-2001; 2001WO-US12865.
XX
PR 21-APR-2000; 2000US-199047P.
PR 02-JUN-2000; 2000US-208841P.
PR 07-JUL-2000; 2000US-216747P.
XX
XX (CORI-) CORIXA CORP.
XX
PI Skeiky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
DR WPI; 2001-616774/71.
DR N-PSDB; AAS59546.
XX
XX Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris.
XX
XX Example 1; SEQ ID NO 11139; 10699P; English.
XX
CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to

CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA).
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX
SQ Sequence 120 AA;

Query Match 4.3%; Score 7; DB 22; Length 120;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 AGPSGCL 9
|111111|
Db 11 agpsgcl 17

RESULT 14
AAG08634 standard; Protein; 149 AA.
XX
AC AAG08634;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 6249.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
KM termination sequence.
XX
XX Arabidopsis thaliana.
XX
XX EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134376.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138450.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.

PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151530.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.

PR 28-OCT-1999: 99US-0161992.
PR 28-OCT-1999: 99US-0161993.
PR 29-OCT-1999: 99US-0162142.

Query Match 4.3%; Score 7; DB 21; Length 149;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 85 GALLEVC 91
Db 70 gallevc 76

RESULT 15

AAR13304
ID AAR13304 standard; Protein; 151 AA.

AC AAR13304;

DT 14-OCT-1991 (first entry)

DE CFTF 556 del A.

Deletion; mutant; diagnosis; antibodies; drug therapy.

Homo sapiens.

Key Location/Qualifiers

FT Domain

FT Domain 81..102
/label= membrane-spanning_domain

FT Domain 118..138
/label= membrane-spanning_domain

FT Modified-site 50
/label= phosphorylation_site

FT Modified-site 63
/note= "by protein kinases C"

FT Modified-site
/label= phosphorylation_site
/note= "by protein kinases C"

PN WO9110734-A.

PD 25-JUL-1991.

PF 11-JAN-1991; 91WO-CA00009.

PR 10-JUL-1990; 90CA-2020817.

PR 12-JAN-1990; 90CA-2007699.

PR 01-MAR-1990; 90CA-2011253.

PA (HSCR-) HSC RES DEV CORP.

PI Tsui LC, Rommens JM, Kerem B;

DR N-PSDB; AAQ13068.

PT Mutant cystic fibrosis trans-membrane conductance regulator gene

PT - used for producing prods. for diagnosis, screening and therapy

PT of cystic fibrosis

PS Clalm 20; Page 124; 178pp; English.

CC 556 del A is a frameshift mutation in exon 4 and results in

CC shortened polypeptides significantly different from the single amino

CC acid deletions or alterations.

CC The mutant CF gene when expressed in cells of the human body, is

CC associated with altered cell function which correlates with the

CC genetic disease cystic fibrosis.

CC See also AAQ13053-72.

Query Match 4.3%; Score 7; DB 12; Length 151;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 155 ISEKLER 161
Db 49 IseKler 55

Search completed: September 18, 2002, 16:36:56
Job time: 291 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2002, 16:35:25 ; Search time 23.9 Seconds

(without alignments)
1172.602 Million cell updates/sec

Title: us-09-676-718a-1

Perfect score: 162
Sequence: 1 MAAGPSGCLVPAPFGKRLLLA.....ILKWNDSVEEPLSEKLERI 162

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 562222 seqs, 172994929 residues

Word size: 0

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

1: SP archaea: *
2: SP bacteria: *
3: SP fungi: *
4: SP human: *
5: SP invertebrate: *
6: SP mammal: *
7: SP mhc: *
8: SP organelle: *
9: SP phage: *
10: SP plant: *
11: SP rodent: *
12: SP virus: *
13: SP vertebrate: *
14: SP unclassified: *
15: SP virus: *
16: SP bacteriophage: *
17: SP archaea: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	146	90.1	162	4	09GZM0	09gzv0 homo sapien
2	95	58.6	162	11	09ERR7	09err7 mus musculus
3	78	48.1	162	11	0923V8	0923v8 rattus norv
4	77	47.5	92	4	09BS64	09bs64 homo sapien
5	9	5.6	152	5	09N4C6	09n4c6 caenorhabd
6	8	4.9	97	16	09A1M1	09a1m1 streptococc
7	8	4.9	142	5	09VVJ7	09vvj7 drosophila
8	4.9	295	16	09S578	09s578 pseudomonas	
9	8	4.9	365	11	09JKU7	09jku7 mus musculus
10	8	4.9	641	2	09WX13	09wx13 brevitacter
11	7	4.3	141	17	096YR9	096yr9 sulfobolus
12	7	4.3	158	10	09SYL3	09syl3 archibopsals
13	7	4.3	159	16	092NX6	092nx6 rhlzobium m
14	7	4.3	173	8	09XKN3	09xkn3 anthophora
15	7	4.3	179	10	09MA71	09ma71 glycine max
16	7	4.3	181	17	027963	027963 archaeoglob

17	7	4.3	196	10	081421
18	7	4.3	197	16	092L50
19	7	4.3	217	13	09DDT8
20	7	4.3	227	10	09LEI4
21	7	4.3	229	10	09LEI5
22	7	4.3	229	10	0948V4
23	7	4.3	233	10	09AVU1
24	7	4.3	238	5	019609
25	7	4.3	238	10	081258
26	7	4.3	240	2	P94207
27	7	4.3	276	16	092J54
28	7	4.3	298	16	09KCS6
29	7	4.3	313	16	0929F3
30	7	4.3	319	10	09M2J6
31	7	4.3	329	10	09SUT6
32	7	4.3	341	10	09FKO5
33	7	4.3	352	10	09S1J9
34	7	4.3	363	10	09L148
35	7	4.3	382	3	09P5J2
36	7	4.3	391	16	09A2D4
37	7	4.3	397	5	094532
38	7	4.3	404	13	09DFB0
39	7	4.3	424	11	09JMB0
40	7	4.3	425	5	09S86
41	7	4.3	426	5	09VLK0
42	7	4.3	440	5	09XX21
43	7	4.3	464	5	09W265
44	7	4.3	480	10	094CK0
45	7	4.3	491	10	09FXK5

ALIGNMENTS

RESULT 1
ID 09GZM0 PRELIMINARY; PRT; 162 AA.
AC 09GZM0;
DT 01-MAR-2001 (TRENBLREL. 16, Created)
DI 01-MAR-2001 (TRENBLREL. 16, last sequence update)
DT 01-MAR-2001 (TRENBLREL. 16, last annotation update)
DE DJ604K5.1 (15 KDA SELENOPROTEIN).
GN DJ604K5.1 OR SEP15.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi.
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid-9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Pearce A.;
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX PubMed-10945981;
RA Kumaraswamy E., Malykh A., Korotkov K.V., Kozayavkin S., Hu Y.,
RA Kwon S.Y., Moustafa M.E., Carlson B.A., Berry M.J., Lee B.J.,
RA Hatfield D.L., Diamond A.M., Gladyshev V.N.;
RT POSSIBLE ROLE OF THE PROTEIN IN CANCER ETIOLOGY.*;
RT "Structure-Expression Relationships of the 15-kDa Selenoprotein Gene.
DR EMBL; AL121989; CAC04186.1; -
DR EMBL; AF288991; AAG31556.1; -
SQ SEQUENCE 162 AA; 17750 MW; 5B94FBB423A089EE CRC64;

Query Match 90.1%; Score 146; DB 4; Length 162;
Best Local Similarity 100.0%; Pred. No. 6.6e-142;
Matches 147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 RLILATVILQAVSAFGAEFSSEACRELGFSNNLTCSSCDLIGQNLQIDPPDCRGCCQEEA 75
DB 16 RLILATVILQAVSAFGAEFSSEACRELGFSNNLTCSSCDLIGQNLQIDPPDCRGCCQEEA 75

Qy	76	QEFENKILVAGAILLEVCGCKKLEFPQVQAFVNSDKPFLRGIOIYVYVGSPPVILKLDDNG	135
Db	76	QEFENKILVAGAILLEVCGCKKLGRRPQVQAFVNSDKRKLFLRGIOIKYVYVGSPPVILKLDDNG	135
Qy	136	NIAEELSLTKNNITDSVEEFLSEKLERI	162
Db	136	NIAEELSLTKNNITDSVEEFLSEKLERI	162

ID	Q9ERR7	PRELIMINARY;	PRT;	162 AA.
AC	Q9ERR7;			
DT	01-MAR-2001 (TREMBLrel, 16, Created)			
DT	01-MAR-2001 (TREMBLrel, 16, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel, 19, Last annotation update)			
DE	15 KDA SELENOPROTEIN.			
CN	SEPI5.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Cranialta; Vertebrata; Euteleostomi;			
CC	Mammalia; Eutheria; Rodentia; Sciurognath; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20519655; Pubmed=10945981;			
RA	Kumaraswamy E., Malikh A., Koroctov K.V., Kozayavkin S., Hu Y.,			
RA	Kwon S.Y., Moustafa M.E., Carlson B.A., Berry M.J., Lee B.J.,			
RA	Haffield D.L., Diamond A.M., Gladyshev V.N.;			
RT	"Structure-Expression Relationships of the 15-kDa Selenoprotein Gene.			
RT	POSIBLE ROLE OF THE PROTEIN IN CANCER ETIOLOGY.";			
RL	J. Biol. Chem. 275:35540-35547(2000).			
DR	EMBL: AF288740; AAC31765.1; "			
DR	MED: MGI:1927947; Sep15.			
SO	SEQUENCE 162 AA; 17738 MW; 6AB4B37D8C8B7847 CRC64;			

[illegible]

RESULT	3		
0923VB			
ID	0923VB	PRELIMINARY;	PRT; 162 AA.
AC	0923VB		
DT	01-DEC-2001 (TrEMBLrel. 19, Created)		
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)		
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)		
DE	15 KDA CYTOSOLIC SELENOPROTEIN.		
OS	Rattus norvegicus (Rat).		
CC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
CC	Mammalia; Eutheria; Rodentia; Scuriognathi; Muridae; Murinae; Rattus.		
OK	NCBI_TaxID=10116;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Roethlein D., Kyriakopoulos A., Behne D.;		
RT	"A 15 kDa-selenoprotein in several tissues of the rat."		
RL	Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AF305054; AAK73100.1; -		
SO	SEQUENCE	162 AA; 17706 MW; 65E939968CB72AC CRC64;	

Query Match	48.1%;	Score 78;	DB 11;	Length 162;
Best Local Similarity	100.0%;	Pred. No. 5.2e-72;		
Matches 79;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy	67	CRGCCCAEAFETTKLYGALILEVCGXKLGFPVQVAFVSDRKLFRGLQIKVRS	SDP	126
Db	67	CRGCCCAEAFETTKLYGALILEVCGXKLGFPVQVAFVSDRKLFRGLQIKVRS	SDP	126
Qy	67	CRGCCCAEAFETTKLYGALILEVCGXKLGFPVQVAFVSDRKLFRGLQIKVRS	SDP	126
Db	67	CRGCCCAEAFETTKLYGALILEVCGXKLGFPVQVAFVSDRKLFRGLQIKVRS	SDP	126
Qy	127	VKLDDNGNIAEELSLK	145	
Db	127	VKLDDNGNIAEELSLK	145	

Q9BS64	PRELIMINARY;	PRF;	92 AA.
ID	Q9BS64.		
AC	Q9BS64.		
DT	01-JUN-2001 (TREMBLrel. 17, Created)		
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)		
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)		
DE	HYPOTHETICAL 9.7 KDA PROTEIN.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=RENAL ADENOCARCINOMA;		
RA	Strausberg R.;		
RL	Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=BONE MARROW, CHRONIC MYELOGENOUS LEUKEMIA;		
RA	Strausberg R.;		
RL	Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL: BC005294; AAH05294.1; -		
DR	EMBL: BC016359; AAH16359.1; -		
KW	Hypothetical protein.		
SQ	SEQUENCE 92 AA; 9669 MW; 59EABAD57FPAAD58 CRC64;		

Query Match	Score	DB	Length
Best Local Similarity	100.0%	Pred. No. 3,36-71	
Matches	77	Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	16	RLILITVIOAVAFAPAEFSEACRELGFSNNLSCSDILGFINLLQDPDRCGCCQGEA	75
DB	16	RLILITVIOAVAFAPAEFSEACRELGFSNNLSCSDILGFINLLQDPDRCGCCQGEA	75
QY	76	QEFKKLLAGALLEVCG	92
DB	76	QEFKKLLAGALLEVCG	92

	RESULT	
GNAC6	5	
09NAC6		
AC	09NAC6	PRELIMINARY; PRT; 152 AA.
DT	01-OCT-2000 (TREMBLrel. 15, Created)	
DT	01-OCT-2001 (TREMBLrel. 18, Last sequence update)	
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)	
DE	HYPOTHEICAL 17.6 KDA PROTEIN.	
GN	V76B12C.3	
OS	Caenorhabditis elegans.	
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;	
OC	Rhabditidae; Pelodierinae; Caenorhabditis.	
OX	NCBI_Taxid=6239;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN-BRISTOL N2;	
RX	MEDLINE=99069613; PubMed=9851916;	
RA	None;	
RT	"Genome sequence of the nematode C. elegans: a platform for	
RT	investigating biology. The C. elegans Sequencing Consortium."	
RL	Science 282:2012-2016(1998).	
RN	[2]	


```

RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Cordes M.;
RT "The sequence of C. elegans cosmid Y76B12C.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC024211; AAF36064.2;
KW Hypothetical protein.
SQ SEQUENCE 152 AA; 17593 MW; 9CAA2P9C0ABE5B87 CRC64;

Query Match          5.6%; Score 9; DB 5; Length 152;
Best Local Similarity 100.0%; Pred. No. 0.41;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 97 REPVOQAFV 105
    |||||
DB 88 REPVOQAFV 96

RESULT 6
O9A1M1 PRELIMINARY; PRT; 97 AA.
AC O9A1M1;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE HYPOTHEICAL PROTEIN SPY0195.
GN SPY0195.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SF370 / ATCC 700294 / SEROTYPE M1;
RX MEDLINE=21192684; PubMed=11296296;
RA Perrella J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
DR EMBL: AE006486; AK33292.1;
DR InterPro: IPR002514; Transposase_8;
DR Pfam: PF01527; Transposase_8;
KM Hypothetical protein: Complete proteome.
SQ SEQUENCE 97 AA; 11418 MW; 81F2B17A7D2B60BB CRC64;

Query Match          4.9%; Score 8; DB 16; Length 97;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 75 AOFETKKL 82
    |||||
DB 64 AOFETKKL 71

RESULT 7
O9VUJ7 PRELIMINARY; PRT; 142 AA.
AC O9VUJ7;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DE 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE CG7484 PROTEIN.
GN CG7484.

```

```

OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt J.G., Nelson C.R., Milos G.L.G.,
RA Abrell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borrova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier A., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasto P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan W., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003523; AAF49314.1;
DR Flybase: FBgn0036745; CG7484.
SQ SEQUENCE 142 AA; 15965 MW; 01D06725BB3DB4DF CRC64;

Query Match          4.9%; Score 8; DB 5; Length 142;
Best Local Similarity 100.0%; Pred. No. 4.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 116 LQIKYVRG 123
    |||||
DB 69 LQIKYVRG 76

RESULT 8
O9S578 PRELIMINARY; PRT; 295 AA.
AC O9S578;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE JR2 PROTEIN (PROBABLE BACTERIOPHAGE PROTEIN).
GN JR2 OR PA0618.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.

```

OX NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 15692 / PAOI;
 RA Nakayama K., Takashima K., Ishihara H., Shinomiya T., Kageyama M.,
 RA Kanaya S., Ohnishi M., Murata T., Terawaki Y., Mori H., Hayashi T.,
 RT "Genetic relationship between bacterioclins and bacteriophages.",
 RL Submitted (Aug-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 15692 / PAOI;
 RA MEDLINE-20437337; PubMed=10984043;
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Goltzy L., Tolelino E., Westbrook-Wadman S., Yuan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Laidig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.,
 RT "Complete genome sequence of *Pseudomonas aeruginosa* PAOI, an
 RT opportunistic pathogen";
 RL Nature 406:959-964(2000).
 DR EMBL; AB030825; BAA83156.1; -;
 DR EMBL; AE004498; AAC04007.1; -;
 KM Complete proteome.
 SQ SEQUENCE 295 AA; 31958 MW; 44444BC3D92CDFE CRC64;

Query Match 4.9%; Score 8; DB 16; Length 295;
 Best Local Similarity 100.0%; Pred. No. 7.9;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 124 SDPVKLL 131
 |||||
 DB 44 SDPVKLL 51

RESULT 9
 O9UKU7 PRELIMINARY; PRT; 365 AA.
 AC O9UKU7;
 DT 01-OCR-2000 (TREMBlrel. 15, Created)
 DT 01-OCR-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-OCR-2000 (TREMBlrel. 15, Last annotation update)
 DE HYPOTHETICAL 40.2 KDA PROTEIN (FRAGMENT).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-C57BL/6J;
 RA MEDLINE-2015471; PubMed=10693804;
 RA Millonig J.H., Millen K.J., Hatten M.E.,
 RT "The mouse Dreher gene *lmx1a* controls formation of the roof plate in
 RT the vertebrate CNS";
 RL Nature 403:764-769(2000).
 DR EMBL; AF226663; AAF43013.1; -;
 KW Hypothetical protein.
 FT NON_TER 1
 SQ SEQUENCE 365 AA; 40209 MW; 86329D3A8BA5C59F CRC64;

Query Match 4.9%; Score 8; DB 11; Length 365;
 Best Local Similarity 100.0%; Pred. No. 9.5;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 20 ATVLQAVS 27
 |||||
 DB 180 ATVLQAVS 187

RESULT 10
 O9WKX13

ID O9WKX13 PRELIMINARY; PRT; 641 AA.
 AC O9WKX13;
 DT 01-NOV-1999 (TREMBlrel. 12, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
 DE ISOMALTOTRIO-DEXTRANASE PRECURSOR PRECURSOR (EC 3.2.1.95).
 GN DEXT.
 OS Brevibacterium fuscum var. dextranolyticum.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteriales; Brevibacterium.
 OC Actinomycetales; Micrococcales; Brevibacteriaceae; Brevibacterium.
 OX NCBI_TaxID=90748;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=0407;
 RA Mizuno T., Mori H., Ito H., Matsui H., Kimura A., Chiba S.,
 RT "Molecular cloning of isomaltootrio-dextranase gene from *Brevibacterium*
 RT fuscum. dextranolyticum strain 0407 and its expression in
 RT *Escherichia coli*";
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB025195; BAA76382.1; -;
 KW Signal; Hydrolase; Glycosidase.
 FT SIGNAL 1 37 POTENTIAL.
 FT CHAIN 38 641 ISOMALTOTRIO-DEXTRANASE.
 SQ SEQUENCE 641 AA; 71988 MW; 27D25EE1AA57CA51 CRC64;

Query Match 4.9%; Score 8; DB 2; Length 641;
 Best Local Similarity 100.0%; Pred. No. 16;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 30 GAFFSSA 37
 |||||
 DB 119 GAFFSSA 126

RESULT 11
 O96YR9 PRELIMINARY; PRT; 141 AA.
 AC O96YR9;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE HYPOTHETICAL PROTEIN S72104.
 GN S72104.
 OS Sulfolobus tokodaii.
 OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
 OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
 OX NCBI_TaxID=111955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-JCM 10545 / 7;
 RX PubMed=11572479;
 RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
 RA Sekine M., Baba S.-I., Anai A., Kosugi H., Hosoyama A., Fukui S.,
 RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
 RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kusuda N., Oguchi A.,
 RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yanagishi A.,
 RA Oshima T., Kikuchi H.,
 RT "Complete genome sequence of an aerobic thermocacidophilic
 RT Crenarchaeon, *Sulfolobus tokodaii* strain7.",
 RL DNA Res. 8:123-140(2001).
 DR EMBL; AP000988; BAB67208.1; -;
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 141 AA; 16616 MW; 000C93003D3EE04D CRC64;

Query Match 4.3%; Score 7; DB 17; Length 141;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 154 FLSEKLE 160
 |||||
 DB 26 FLSEKLE 32

```

RESULT 12
Q9SYL3 PRELIMINARY; PRT; 158 AA.
AC Q9SYL3;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE F3F20.17 PROTEIN.
GN F3F20.17
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Federspiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,
RA Altif H., Araujo R., Huizar L., Rowley D., Buehler E., Dunn P.,
RA Gonzalez A., Kremenetskaia I., Kim C., Lenz C., Li J., Liu S.,
RA Lucos S., Schwartz J., Shim P., Toriumi M., Vysotskaya V.S.,
RA Walker M., Yu G., Ecker J., Theologis A., Davis R.W.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC007153; AAD30621.1; -
DR InterPro; IPR000345; CYC_heme_bind.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN.1.
SQ SEQUENCE 158 AA; 17901 MW; A181CB37DA433B8B CRC64;

Query Match 4.3%; Score 7; DB 10; Length 158;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 85 GALEVC 91
|111111|
DB 79 GALEVC 85

RESULT 13
Q92NX6 PRELIMINARY; PRT; 159 AA.
AC Q92NX6;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE PUTATIVE RNA POLYMERASE SIGMA FACTOR PFCI PROTEIN.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RA MEDLINE=21368334; PubMed=11474104;
RA Gallibert F., Finan T.M., Long S.R., Puehler A., Abola P., Ampe F.,
RA Barloy-Hubler F., Barnett M.J., Becker A., Boisdard P., Bothe G.,
RA Boutry M., Bowser L., Buhrmester J., Cadieu E., Capela D., Chain P.,
RA Cowie A., Davis R.W., Dreano S., Federspiel N.A., Fisher R.F.,
RA Gloux S., Godle T., Goffeau A., Golding B., Gonzy J., Gurjal M.,
RA Hernandez-Lucas I., Hong A., Huizar L., Hyman R.W., Jones T., Kahn D.,
RA Kahn M.L., Kalman S., Keating D.H., Kiss E., Komp C., Lelaure V.,
RA Masny D., Palm C., Peck M.C., Pohl T.M., Portetlelle D., Purnelle B.,
RA Rampeger U., Surzycki R., Thebaud P., Vandebol M.,
RA Vorhoeelter F.J., Weidner S., Wells D.H., Wong K., Yeh K.-C., Batut J.;
RL "The composite genome of the legume symbiont Sinorhizobium meliloti.";
RT Science 293:668-672(2001).
DR EMBL; AL591789; CAC4610.1; -
KW Complete proteome.
SQ SEQUENCE 159 AA; 18395 MW; 5E3A1654D55C3010 CRC64;

Query Match 4.3%; Score 7; DB 16; Length 159;

```

```

Best Local Similarity 100.0%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 137 IAEELSI 143
|111111|
DB 129 IAEELSI 135

RESULT 14
Q9XKN3 PRELIMINARY; PRT; 173 AA.
AC Q9XKN3;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE CYTOCHROME B (FRAGMENT).
GN CYTB.
OS Anthophora acervorum.
OC Mitochondrion.
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata;
OC Apoidea; Apidae; Anthophora.
OX NCBI_TaxID=60897;
RN [1]
RP SEQUENCE FROM N.A.
RA Koulanas S., Schmid-Hempel R., Roubik D., Schmid-Hempel P.;
RT "Relationships within the Apinae."
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC COUPLED TO ATP SYNTHESIS (BY SIMILARITY).
CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
CC BOUND TO THE PROTEIN (BY SIMILARITY).
CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CC CYTOCHROME C1 AND THE RIESKE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.
DR EMBL; AF002731; AAD29081.1; -
DR InterPro; IPR000179; Cyt_b_b6.
DR Pfam; PF00033; cytochrome_b_N; 1.
KW Electron transport; Heme; Mitochondrion; Respiratory chain;
KW Transmembrane.
FT NON_TER 1
FT NON_TER 173
SQ SEQUENCE 173 AA; 20314 MW; C2BA74FEA9F40FC1 CRC64;

Query Match 4.3%; Score 7; DB 8; Length 173;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 41 LGFSSNL 47
|111111|
DB 121 LGFSSNL 127

RESULT 15
Q9MAZ1 PRELIMINARY; PRT; 179 AA.
AC Q9MAZ1;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE NONCATALYTIC COAT PROTEIN ZETA2-COP.
GN COPZ2.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX NCBI_TaxID=3847;
RN [1]
RP SEQUENCE FROM N.A.
RA Hahn Y., Chung J.H.;

```

RT "Identification of zeta-COP genes from various organisms."
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AB040543; BAA93004.1; -
 DR InterPro; IPR000804; Clat_adaptor_s.
 DR Pfam; PF01217; Clat_adaptor_s; 1.
 KW Coat protein.
 SQ SEQUENCE 179 AA; 19790 MW; CC01CFE2EB84341C CRC64;

Query Match 4.3%; Score 7; DB 10; Length 179;
 Best Local Similarity 100.0%; Pred. No. 54;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 19 LATVLA 25
 |||||
 Db 90 LATVLA 96

Search completed: September 18, 2002, 16:38:58
 Job time: 213 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2002, 16:33:55 ; Search time 18.81 Seconds
(without alignments)
827.564 Million cell updates/sec

Title: US-09-676-718a-1

Perfect score: 162

Sequence: 1 MAAGPSCGLVPAFGKRLLLA.....ILKMTDSVEEFLSEKLERI 162

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283138 seqs, 96089334 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_71:*

1: pir1:*\n2: pir2:*\n3: pir3:*\n4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	4.9	295	2	T44541
2	8	4.9	641	2	JC7142
3	8	4.9	743	2	E83728
4	8	4.9	878	2	AB0478
5	7	4.3	83	1	S25307
6	7	4.3	145	2	S10027
7	7	4.3	158	2	E86191
8	7	4.3	173	2	S10975
9	7	4.3	181	2	A69540
10	7	4.3	196	2	T01425
11	7	4.3	197	2	T26829
12	7	4.3	197	2	C88809
13	7	4.3	282	2	AE2618
14	7	4.3	290	2	D97400
15	7	4.3	298	2	E83836
16	7	4.3	313	2	AG1722
17	7	4.3	319	2	T15980
18	7	4.3	320	2	S32966
19	7	4.3	329	2	T13016
20	7	4.3	352	2	C84603
21	7	4.3	363	2	T49209
22	7	4.3	365	1	MNXRAH
23	7	4.3	382	2	T49724
24	7	4.3	390	1	AA2724
25	7	4.3	390	2	AG2463
26	7	4.3	391	2	E87699
27	7	4.3	396	2	H82728
28	7	4.3	412	1	ITMSC
29	7	4.3	424	2	A50000

30	7	4.3	455	2	T26730	hypothetical prote
31	7	4.3	469	2	G65058	hypothetical prote
32	7	4.3	478	2	D64230	NADH oxidase (nox)
33	7	4.3	491	2	B86155	probable polygalac
34	7	4.3	505	2	T35107	hypothetical prote
35	7	4.3	507	2	B69957	conserved hypothet
36	7	4.3	510	2	A53802	N-acetylglactosam
37	7	4.3	512	2	AF3402	exopolysphatase
38	7	4.3	516	2	T33061	hypothetical prote
39	7	4.3	600	2	B84546	probable splicing
40	7	4.3	627	2	D71729	dnar-type molecula
41	7	4.3	627	2	A97729	dnar protein (limpo
42	7	4.3	660	2	S71949	metalloproteinase
43	7	4.3	667	2	S74254	homeotic protein s
44	7	4.3	674	2	T40214	hypothetical prote
45	7	4.3	770	2	T01527	hypothetical prote

ALIGNMENTS

RESULT 1
T44541
hypothetical protein JR2 [imported] - Pseudomonas aeruginosa
N:Alternate names: probable bacteriophage protein PA0618
C:Species: Pseudomonas aeruginosa
C:Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 31-Dec-2000
C:Accession: T44541; H83566
R:Nakayama, K.; Takashima, K.; Ishihara, H.; Shinomiya, T.; Kageyama, M.; Kanaya, S.;
submitted to the EMBL Data Library, August 1999
A:Description: Genetic relationship between bacteriophages and bacteriophages.
A:Reference number: 222790
A:Accession: T44541
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-295 <NAK>
A:Cross-References: EMBL:AB030825; PIDN:BAH83156.1
A:Experimental source: strain PA01
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lapid, K.; L
; Lory, S.; Olson, M.V.
N:ature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Reference number: AB2950; MUID:20437337
A:Accession: H83566
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-295 <STO>
A:Cross-References: GB:AE004498; GB:AE004091; NID:99946491; PIDN:AA04007.1; GSPDB:GN
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA0618
A>Note: JR2

Query Match 4.9%; Score 8; DB 2; Length 295;
Best Local Similarity 100.0%; Pred. No. 3.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 124 SDPVKLL 131
DB 44 SDPVKLL 51

RESULT 2
JC7142
dextranase (EC 3.2.1.11) - Brevibacterium fuscum var. dextranolyticum
N:Alternate names: isomaltotriose-dextranase
C:Species: Brevibacterium fuscum var. dextranolyticum
C:Date: 04-Mar-2000 #sequence_revision 04-Mar-2000 #text_change 11-May-2000
C:Accession: JC7142; PC7042
R:Mizuno, T.; Mori, H.; Ito, H.; Matsui, H.; Kimura, A.; Chiba, S.
Biosci. Biotechnol. Biochem. 63, 1582-1588, 1999

A:Title: Molecular cloning of isomaltotrio-dextranase gene from *Brevibacterium fuscum* v4
 A:Reference number: JCI7142; MUID:20008261
 A:Accession: JCI7142
 A:Molecule type: DNA
 A:Residues: 1-641 <MIZ>
 A:Cross-references: DDBJ:AB025195
 A:Experimental source: strain 0407
 A:Accession: PC7042
 A:Molecule type: protein
 A:Residues: 38-56;98-109;141-147;149-169;213-231;298-307;309-321;323-342;446-458;497-515
 C:Genetics:
 A:Gene: dext
 C:Keywords: glycosidase; hydrolase

Query Match 4.9%; Score 8; DB 2; Length 641;
 Best Local Similarity 100.0%; Pred. No. 7.4;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 30 GAFFSEA 37
 DB 119 GAFFSEA 126

RESULT 3
 E83728
 Phosphoribosylformylglycinamide synthetase I puro [imported] - *Bacillus halodurans* (st
 C:Species: *Bacillus halodurans*
 C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
 C:Accession: E83728
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hird
 Nucleic Acids Res. 28, 4317-4331, 2000
 A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and
 A:Reference number: A83650; MUID:20512582; PMID:11058132
 A:Accession: E83728
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-743 <STO>
 A:Cross-references: GB:AP001509; GB:BA000004; NID:g10173176; PIDN:BAB04348.1; GSPDB:GN00
 A:Experimental source: strain C-125
 C:Genetics:
 A:Gene: puro
 C:Superfamily: phosphoribosylformylglycinamide synthase component II

Query Match 4.9%; Score 8; DB 2; Length 743;
 Best Local Similarity 100.0%; Pred. No. 8.4;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 35 SEACRELG 42
 DB 525 SEACRELG 532

RESULT 4
 AE0478
 Phosphoenolpyruvate carboxylase (EC 4.1.1.31) [imported] - *Yersinia pestis* (strain CO92)
 C:Species: *Yersinia pestis*
 C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 27-Nov-2001
 C:Accession: AE0478
 R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tittball, R.W.; Holden, M.T.G.; Prentice, M.B.
 deno-Tharaga, A.M.; Chillingworth, T.; Croft, A.; Davies, R.M.; Davis, P.; Dougan, G.;
 I. M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall,
 Nature 413, 523-527, 2001
 A:Title: Genome sequence of *Yersinia pestis*, the causative agent of plague.
 A:Reference number: AB00001; MUID:21470413; PMID:11586360
 A:Accession: AE0478
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-878 <RUR>
 A:Cross-references: GB:AL590842; PIDN:CAC93393.1; PID:g15981839; GSPDB:GN00175
 C:Genetics:
 A:Gene: ppc

C:Superfamily: phosphoenolpyruvate carboxylase
 C:Keywords: carbon-carbon lyase; carboxy-lyase

Query Match 4.9%; Score 8; DB 2; Length 878;
 Best Local Similarity 100.0%; Pred. No. 9.7;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 82 LYAGALE 89
 DB 632 LYAGALE 639

RESULT 5
 S25307
 Probable allophycocyanin linker protein - red alga (*Cyanidium caldarium*) chloroplast
 C:Species: chloroplast *Cyanidium caldarium*
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C:Accession: S25307
 R:Valentin, K.; Mäld, U.; Emlich, A.; Zetsche, K.
 Plant Mol. Biol. 20, 267-276, 1992
 A:Title: Organization and expression of a phycobilliprotein gene cluster from the unic
 A:Reference number: S25306; MUID:93004479
 A:Accession: S25307
 A:Molecule type: DNA
 A:Residues: 1-83 <VAL>
 A:Cross-references: EMBL:X57251; NID:g17969; PIDN:CAA40532.1; PID:g17971
 C:Genetics:
 A:Gene: apcL
 A:Superfamily: chloroplast
 C:Keywords: chloroplast

Query Match 4.3%; Score 7; DB 1; Length 83;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 43 FSSNMLC 49
 DB 13 FSSNMLC 19

RESULT 6
 S10027
 alpha-amylase inhibitor, tetrameric, chain CMI precursor - wheat
 C:Species: *Triticum aestivum* (common wheat)
 C:Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 20-Jun-2000
 C:Accession: S10027
 R:Garcia-Maroto, F.; Marana, C.; Mena, M.; Garcia-Olmedo, F.; Carbonero, F.
 Plant Mol. Biol. 14, 845-853, 1990
 A:Title: Cloning of cDNA and chromosomal location of genes encoding the three types o
 A:Reference number: S10027; MUID:91346675
 A:Accession: S10027
 A:Molecule type: mRNA
 A:Residues: 1-145 <GAR>
 A:Cross-references: EMBL:X17575; NID:g21700; PIDN:CAA35598.1; PID:g21701
 C:Superfamily: wheat alpha-amylase inhibitor
 C:Keywords: alpha-amylase inhibitor; tetramer
 F:1-25/Domain: signal sequence #status predicted <SIG>
 F:26-145/Product: alpha-amylase inhibitor chain CMI #status predicted <MAT>

Query Match 4.3%; Score 7; DB 2; Length 145;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 17 LLATVL 23
 DB 10 LLATVL 16

RESULT 7

E86191
 hypothetical protein [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: E86191
 R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federapfel, N.A.; Kaul, S.; White, O.; Alonso, Chih, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Hutzar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.A.; Li, J.H.; Lin, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Matli, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719
 A:Accession: E86191
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-158 <STO>
 A:Cross-references: GB:AEO05172; NID:G4836919; PID:AU030621.1; GSPDB:GN00141
 C:Genetics:
 A:Map position: 1

Query Match 4.3%; Score 7; DB 2; Length 158;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 85 GATLEVC 91
 |||||
 DB 79 GATLEVC 85

RESULT 8
 S10975
 T-cell surface glycoprotein CD3 delta chain precursor - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 18-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 21-Jan-2000
 C:Accession: S10975
 R: Davies, J.D.; Mueller, D.; Wilson, D.B.; Gold, D.P.
 Nucleic Acids Res. 18, 4617, 1990
 A:Title: Nucleotide sequence of a cDNA encoding the rat T3 delta chain.
 A:Reference number: S10975; MUID:90356424
 A:Accession: S10975
 A:Molecule type: mRNA
 A:Residues: 1-173 <DAV>
 A:Cross-references: EMBL:X53430; NID:G55913; PID:CAA37521.1; PID:G55914
 C:Superfamily: T-cell surface glycoprotein CD3 delta chain; Immunoglobulin homology
 C:Keywords: glycoprotein; transmembrane protein
 F:30-75/Domain: Immunoglobulin homology <IMM>

Query Match 4.3%; Score 7; DB 2; Length 173;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 17 LLLATVL 23
 |||||
 DB 10 LLLATVL 16

RESULT 9
 A69540
 conserved hypothetical protein AF2321 - Archaeoglobus fulgidus
 C:Species: Archaeoglobus fulgidus
 C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999
 C:Accession: A69540
 R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson, J.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L. Nature 390, 364-370, 1997
 A:Authors: Ullrich, T.; Cotton, M.D.; Spriggs, T.; Atliach, P.; Kaine, B.P.; Sykes, S.

Smith, H.O.; Woese, C.R.; Venter, J.C.
 A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch
 A:Reference number: A69540; MUID:98049343
 A:Accession: A69540
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-181 <KLE>
 A:Cross-references: GB:AEO00944; GB:AEO00782; NID:G2689267; PID:AB88935.1; PID:G264

Query Match 4.3%; Score 7; DB 2; Length 181;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 35 SEACREL 41
 |||||
 DB 77 SEACREL 83

RESULT 10
 T01425
 hypothetical protein T2H3.15 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 24-Mar-1999
 C:Accession: T01425
 R:Kalicki, J.; Gibson, A.
 Submitted to the EMBL Data Library, August 1998
 A:Description: The sequence of A. thaliana T2H3.
 A:Reference number: Z14324
 A:Accession: T01425
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-196 <KAL>
 A:Cross-references: EMBL:AF075597; NID:G3298610; PID:G3377809
 A:Experimental source: cultivar Columbia
 C:Genetics:
 A:Map position: 4
 A:Introns: 77/1
 A:Note: T2H3.15

Query Match 4.3%; Score 7; DB 2; Length 196;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 138 AEBLSIL 144
 |||||
 DB 29 AEBLSIL 35

RESULT 11
 T26829
 hypothetical protein Y43C5B.1 - Caenorhabditis elegans (fragment)
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T26829
 R:White, S.
 submitted to the EMBL Data Library, June 1998
 A:Reference number: Z20273
 A:Accession: T26829
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-197 <WIL>
 A:Cross-references: EMBL:AL023841; PID:CAA19512.1; GSPDB:GN00022; CESP:Y43C5B.1
 A:Experimental source: clone Y43C5B
 C:Genetics:
 A:Gene: CESP:Y43C5B.1
 A:Map position: 4

Query Match 4.3%; Score 7; DB 2; Length 197;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 FSSBACR 39
 |||||
 DB 18 FSSBACR 24

RESULT 12

C88809
 Protein V43C5B.1 [Imported] - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
 C:Accession: C88809
 R:Anonymous, The C. elegans Sequencing Consortium.
 Science 282, 2012-2018, 1998

A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
 A:Reference number: A75000; MUID:99069613; PMID:9851916
 A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/projects/C_elegans/
 A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
 A:Accession: C88809
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-197 <STO>
 A:Cross-references: GB:chr_IV; PIDN:CA19512.1; PID:g380944; GSPDB:GN00022; CESP:V43C5B
 C:Genetics:
 A:Gene: V43C5B.1
 A:Map position: 4

Query Match 4.3%; Score 7; DB 2; Length 197;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 FSSBACR 39
 |||||
 DB 18 FSSBACR 24

RESULT 13

AE2618
 rRNA methylase [Imported] - Agrobacterium tumefaciens (strain C58, Dupont)
 C:Species: Agrobacterium tumefaciens
 C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 01-Feb-2002
 C:Accession: AE2618
 R:Mod, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.
 erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kuyavirin, T.; Levy, R.; Li, M.; McClellan,
 ; Karp, P.; Romero, P.; Zhang, S.
 Science 294, 2317-2323, 2001
 A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
 ster, E.W.
 A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
 A:Reference number: AB2577; PMID:11743193
 A:Accession: AE2618
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-282 <KUR>
 A:Cross-references: GB:AE008688; PIDN:AA141363.1; PID:g17738678; GSPDB:GN00186
 A:Experimental source: strain C58 (Dupont)
 C:Genetics:
 A:Gene: trmH
 A:Map position: circular chromosome
 C:Superfamily: conserved hypothetical protein HI0860

Query Match 4.3%; Score 7; DB 2; Length 282;
 Best Local Similarity 100.0%; Pred. No. 39;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 150 SVEEFLS 156
 |||||
 DB 188 SVEEFLS 194

RESULT 14

D97400
 rRNA methylase (AP001517) [Imported] - Agrobacterium tumefaciens (strain C58, Cereon)
 C:Species: Agrobacterium tumefaciens
 C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
 C:Accession: D97400
 R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldm
 A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz,
 Science 294, 2323-2328, 2001
 A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium
 A:Reference number: A97359; PMID:11743194
 A:Accession: D97400
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-290 <KUR>
 A:Cross-references: GB:AE007869; PIDN:AAK86157.1; PID:g15155248; GSPDB:GN00169
 C:Genetics:
 A:Gene: AGR_C_593
 A:Map position: circular chromosome
 C:Superfamily: conserved hypothetical protein HI0860

Query Match 4.3%; Score 7; DB 2; Length 290;
 Best Local Similarity 100.0%; Pred. No. 40;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 150 SVEEFLS 156
 |||||
 DB 196 SVEEFLS 202

RESULT 15

E83836
 reductase BH1493 [Imported] - Bacillus halodurans (strain C-125)
 C:Species: Bacillus halodurans
 C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
 C:Accession: E83836
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H
 Nucleic Acids Res. 28, 4317-4331, 2000
 A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a
 A:Reference number: A83650; MUID:20512582; PMID:11058132
 A:Accession: E83836
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-298 <STO>
 A:Cross-references: GB:AP001512; GB:BA000004; MID:g10174030; PIDN:BA05212.1; GSPDB:G
 A:Experimental source: strain C-125
 C:Genetics:
 A:Gene: BH1493

Query Match 4.3%; Score 7; DB 2; Length 298;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 KRLLIAT 21
 |||||
 DB 100 KRLLIAT 106

Search completed: September 18, 2002, 16:38:13
 Job time: 258 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2002, 16:36:11 ; Search time 10.7 Seconds

(without alignments)
586.221 Million cell updates/sec

Title: US-09-676-718A-1

Perfect score: 162
Sequence: 1 MAAGPSGCLVPAFGKRLLLA.....ILKMWNDSVEEFLSEKLERI 162

Scoring table:
OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 105224 seqs, 38719550 residues

Word size : 0

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	77	47.5	162	SE15_HUMAN	O60613 homo sapien
2	69	42.6	162	SE15_MOUSE	O96877 mus musculu
3	69	42.6	162	SE15_RAT	O923V8 rattus norv
4	8	4.9	640	DEXT_ARTGO	P70744 arthrobacte
5	8	4.9	743	PURL_BACHD	O9F57 bacillus ha
6	7	4.3	83	YC20_GALSD	P48409 tridictoria s
7	7	4.3	145	IA01_WHEAT	P16850 tritlicum ae
8	7	4.3	165	UR1_ONCAT	O93448 oncorhynch
9	7	4.3	173	CB3D_RAT	P19377 rattus norv
10	7	4.3	210	KAB2_OLDAF	P38454 oldenlandia
11	7	4.3	320	YB9K_YEAST	P58452 saccharomyc
12	7	4.3	365	VNS2_AHSV9	P27279 african hor
13	7	4.3	390	RPSD_ANASP	P26683 anabaena sp
14	7	4.3	396	DKR_XYLPA	O96810 xyliella fas
15	7	4.3	412	ALAT_MUSCR	P26595 mus caroli
16	7	4.3	433	HEM2_SPIOL	P24493 spinacia ol
17	7	4.3	445	YGC5_ECOLI	O46909 escherichia
18	7	4.3	478	NAOX_MYGE	O49408 mycoplasma
19	7	4.3	507	YOGP_BAGSU	P54493 bacillus su
20	7	4.3	510	BAGT_MOUSE	O09199 mus musculu
21	7	4.3	627	DNAK_RICCN	O92136 rickettsia
22	7	4.3	627	DNAX_MOUSE	O92493 rickettsia
23	7	4.3	667	SIX5_MOUSE	P70178 mus musculu
24	7	4.3	674	CWF4_SCHPO	P87312 schizosacch
25	7	4.3	830	GYRA_GLOAB	P94605 clostridium
26	7	4.3	831	YPDD_ECOLI	P77439 escherichia
27	7	4.3	1284	NRX4_DROME	O94887 drosophila
28	7	4.3	1450	CFTR_RABIT	O00554 oryctolagus
29	7	4.3	1476	CFTR_MOUSE	P26361 mus musculu
30	7	4.3	1480	CFTR_HUMAN	P15569 homo sapien
31	7	4.3	1481	CFTR_BOVIN	P35071 bos taurus
32	7	4.3	1481	CFTR_SHEEP	O00555 ovins aries
33	6	3.7	79	Y9K_HCMVA	P21600 human cytom

34	6	3.7	85	Y0AH_BACSU	P45905 bacillus su
35	6	3.7	90	REV_HV1ZH	P05868 human immun
36	6	3.7	95	YCNE_BACSU	P94425 bacillus su
37	6	3.7	100	R14_CYPAP	P48138 cyanophora
38	6	3.7	119	R18E_SULSO	P95990 sulfolobus
39	6	3.7	121	SRI4_ARATH	O04421 arabidopsis
40	6	3.7	129	LVC_EQUAS	P11375 equus asinu
41	6	3.7	131	AGSM_MOUSE	O03488 mus musculu
42	6	3.7	131	AGSM_VULVU	P79407 vulpes vulp
43	6	3.7	131	ATPE_GUTH	O78492 guillardi
44	6	3.7	131	SVR_BUCAP	O08888 buchnera ap
45	6	3.7	132	AGSM_HUMAN	P42127 homo sapien

ALIGNMENTS

```

RESULT 1
ID SE15_HUMAN STANDARD; PRT; 162 AA.
AC O60613; O9NR01; O9BS64; O9GZM0;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE 15 kDa selenoprotein precursor.
GN SEPI5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 98-106; 123-127 AND 146-158.
RX MEDLINE-98204881; PubMed-9535873;
RA Gladyshev V.N., Jeang K.-T., Wootton J.C., Hatfield D.L.;
RT "A new human selenium-containing protein. Purification,
RT characterization, and cDNA sequence."
RL J. Biol. Chem. 273:8910-8915(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-20519655; PubMed-10945981;
RA Kumarasamy E., Malyskh A., Korotkov R.V., Kozayavkin S., Hu Y.,
RA Kwon S.Y., Moustafa M.E., Carlson B.A., Berry M.J., Lee B.J.,
RA Hatfield D.L., Diamond A.M., Gladyshev V.N.;
RT "Structure-expression relationships of the 15-kDa selenoprotein gene.
RT Possible role of the protein in cancer etiology."
RL J. Biol. Chem. 275:35540-35547(2000).
RN [3]
RP SEQUENCE FROM N.A.
RA Ryu M., Moon E.;
RT "The human 15-kDa selenoprotein gene: characterisation of the genomic
RT structure and functional analysis of the promoter."
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Pearce A.;
RT Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA TISSUE-Bone marrow, and kidney;
RT Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SUBUNIT: SEEMS TO BE EITHER OLIGOMERIC OR HETEROOLIGOMERIC.
CC -1- PTM: THE N-TERMINUS IS BLOCKED.
CC -1- MASS SPECTROMETRY: MW=14870; METHOD-Electrospray; RANGE=7-152.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).

```

```

DR EMBL: AF051894; AAC15478.1; -
DR EMBL: AF288891; AAC31556.1; -
DR EMBL: AF288992; AAC31557.1; -
DR EMBL: AF267982; AAF78966.1; ALT_SEQ.
DR EMBL: AF267982; AAF78966.1; JOINED.
DR EMBL: AF267983; AAF78966.1; JOINED.
DR EMBL: AF267984; AAF78966.1; JOINED.
DR EMBL: AF267985; AAF78966.1; JOINED.
DR EMBL: AL121989; CAC04186.1; ALT_SEQ.
DR EMBL: BC005294; AAH05294.1; ALT_TERM.
DR EMBL: BC016359; AAH16359.1; ALT_TERM.
DR MIM: 606254; -
KW Selenium; Selenocysteine; Signal.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 162 15 KDA SELENOPROTEIN.
FT SE_CYS 93 93
SQ SEQUENCE 162 AA; 17743 MW; 463EACBA23B4CDBE CRC64;

Query Match 47.5%; Score 77; DB 1; Length 162;
Best Local Similarity 100.0%; Pred. No. 6,3e-71;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 16 RLATATVQAASARGAERSSACRELGFSNLTGSCDILGQFNLTOLDPDCGCCOEAA 75
DB 16 RLATATVQAASARGAERSSACRELGFSNLTGSCDILGQFNLTOLDPDCGCCOEAA 75
OY 76 QFETKRLVAGALIEVCG 92
DB 76 QFETKRLVAGALIEVCG 92

RESULT 2
SEIS_MOUSE STANDARD; PRT; 162 AA.
ID SEIS_MOUSE
AC O9ERR7;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update).
DE 15 kDa selenoprotein precursor.
GN SEPI5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=20519655; PubMed=10945981;
RA Kumaraswamy E., Malykh A., Korotkov K.V., Kozvaykin S., Hu Y.,
RA Kwon S.Y., Moustafa M.E., Carlson B.A., Berry M.J., Lee B.J.,
RA Hatfield D.L., Diamond A.M., Gladyshev V.N.;
RT "Structure-expression relationships of the 15-kDa selenoprotein gene.
RT Possible role of the protein in cancer etiology.";
RT J. Biol. Chem. 275:35540-35547(2000).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF288740; AAC31765.1; ALT_SEQ.
DR MGD; MGI:1927947; Sep15.
KW Selenium; Selenocysteine; Signal.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 162 15 KDA SELENOPROTEIN.
FT SE_CYS 93 93
SQ SEQUENCE 162 AA; 17731 MW; 771E047D8C9F3C17 CRC64;

Query Match 42.6%; Score 69; DB 1; Length 162;

```

```

Best Local Similarity 100.0%; Pred. No. 8,6e-63;
Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 94 KLGRFPQVAVRSDKPLFRGLQIKYVGSDPYLKLLDNGNIAEELSTIKWNTDSYEE 153
DB 94 KLGRFPQVAVRSDKPLFRGLQIKYVGSDPYLKLLDNGNIAEELSTIKWNTDSYEE 153
OY 154 FLSEKLERI 162
DB 154 FLSEKLERI 162

RESULT 3
SEIS_RAT STANDARD; PRT; 162 AA.
ID SEIS_RAT
AC O923V8;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE 15 kDa selenoprotein precursor.
GN SEPI5.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Roethlein D., Kyriakopoulos A., Behne D.;
RT "A 15 kDa-selenoprotein in several tissues of the rat.";
RT Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF390544; AAR73100.1; ALT_SEQ.
KW Selenium; Selenocysteine; Signal.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 162 15 KDA SELENOPROTEIN.
FT SE_CYS 93 93
SQ SEQUENCE 162 AA; 17775 MW; 73985E968C9F36FC CRC64;

Query Match 42.6%; Score 69; DB 1; Length 162;
Best Local Similarity 100.0%; Pred. No. 8,6e-63;
Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 94 KLGRFPQVAVRSDKPLFRGLQIKYVGSDPYLKLLDNGNIAEELSTIKWNTDSYEE 153
DB 94 KLGRFPQVAVRSDKPLFRGLQIKYVGSDPYLKLLDNGNIAEELSTIKWNTDSYEE 153
OY 154 FLSEKLERI 162
DB 154 FLSEKLERI 162

RESULT 4
DEXT_ARTGO STANDARD; PRT; 640 AA.
ID DEXT_ARTGO
AC P70744;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Dextranase precursor (EC 3.2.1.11) (Alpha-1,6-glucan-6-
DE glucanohydrolase) (Endodextrinase).
OS Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Thermoplasma.
OC Actinomyces; Actinobacteria; Actinobacteridae; Actinobacteriales;
OC Actinomycetales; Micrococcales; Micrococcaceae; Arthrobacter.
OX NCBI_TaxID=1665;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-T-3044;
RA Oguma T., Kurokawa T., Tobe K., Kitao S., Kobayashi M.;
RL Submitted (NOV-1996) to the EMBL/Genbank/DDJ databases.
CC -1- CATALYTIC ACTIVITY: Endohydrolays of 1,6-alpha-D-glucosidic
CC linkages in dextran.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO FAMILY 49 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D88361; BAA13596.1; -
KW Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 32
FT CHAIN 33 640
FT DEXTRANASE.
SQ SEQUENCE 640 AA; 71142 MW; 4AC7864A96A537D2 CRC64;

Query Match 4.9%; Score 8; DB 1; Length 640;
Best Local Similarity 100.0%; Pred. No. 3.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 30 GAEFSSEA 37
DB 125 GAEFSSEA 132

RESULT 5
PRRL_BACCHD STANDARD; PRT; 743 AA.
AC 09KE57;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Phosphoribosylformylglycinamide synthase II (EC 6.3.5.3) (FGAM
DE synthase II).
GN PURL OR BH0629.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=8665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fujii F., Hirata C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
CC -1- CATALYTIC ACTIVITY: ATP + 5'-phosphoribosylformylglycinamide + L-
CC glutamine + H(2)O = ADP + phosphate + 5'-
CC phosphoribosylformylglycinamide + L-glutamate.
CC -1- PATHWAY: DE NOVO PURINE BIOSYNTHESIS; FOURTH STEP.
CC -1- SUBUNIT: HETERODIMER OF TWO SUBUNTS; PURO AND PURL.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE FGAMS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC -----
DR EMBL; AP001509; BAB04348.1; -
DR InterPro; IPR000728; AIRS_related.
DR Pfam; PF00586; AIRS; 2.
DR Pfam; PF02769; AIRS_C; 2.
KW Purine biosynthesis; Ligase; ATP-binding; Complete proteome.
FT NP_BIND 111 122
FT ATP (POTENTIAL).
SQ SEQUENCE 743 AA; 80008 MW; 005FC0855D20D84D CRC64;

Query Match 4.9%; Score 8; DB 1; Length 743;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 35 SEACRELG 42
DB 525 SEACRELG 532

RESULT 6
YC20_GALSU STANDARD; PRT; 83 AA.
ID YC20_GALSU
AC P48A09;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 9.5 kDa protein ycf20.
GN YCF20.
OS Caldectria sulphuraria.
OG Chloroplast.
OC Eukaryota; Rhodophyta; Bangiophyceae; Porphyridiales; Porphyridiaceae;
OC Galdieria.
OX NCBI_TaxID=130081;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=14-1-1 / ISOLATE 107.79/GOETTINGEN;
RX MEDLINE=93004479; PubMed=1391770;
RA Valentin K., Maid U., Emich A., Zetsche K.;
RT "Organization and expression of a phycoobiliprotein gene cluster from
RT the unicellular red alga Cyanidium caldarium.";
RL Plant Mol. Biol. 20:267-276(1992).
CC -1- SIMILARITY: BELONGS TO THE YCF20 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X57251; CAA40532.1; -
DR PIR; S25307; S25307.
KW Chloroplast; Hypothetical protein.
SQ SEQUENCE 83 AA; 9485 MW; 990117EC63FE3EF9 CRC64;

Query Match 4.3%; Score 7; DB 1; Length 83;
Best Local Similarity 100.0%; Pred. No. 5.5;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 43 FSSNLIC 49
DB 13 FSSNLIC 19

RESULT 7
IA01_WHEAT STANDARD; PRT; 145 AA.
ID IA01_WHEAT
AC P16850;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-AUG-1990 (Rel. 15, Last annotation update)

```

```

DE Alpha-amylase/trypsin inhibitor CMI precursor (Chloroform/methanol-
DE soluble protein CMI).
OS Triticum aestivum (Wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
OC Triticeae; Triticum.
CX NCBI_TaxID=4565;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, CHINESE SPRING; TISSUE=Endosperm;
RX MEDLINE=91346675; Pubmed=2102861;
RA Garcia-Maroto F., Marana C., Mena M., Garcia-Olmedo F., Carbonero P.;
RT "Cloning of cDNA and chromosomal location of genes encoding the three
RT types of subunits of the wheat tetrameric inhibitor of insect alpha-
RT amylase."
RL Plant Mol. Biol. 14:845-853(1990).
RN [2]
RP SEQUENCE OF 26-53.
RC STRAIN=CV, CANDEL:
RA Barber D., Sanchez-Monge R., Garcia-Olmedo F., Salcedo G., Mendez E.;
RT "Evolutionary implications of sequential homologs among members of
RT the trypsin / alpha-amylase inhibitor family (CM-proteins) in wheat
RT and barley."
RL Biochim. Biophys. Acta 873:147-151(1986).
CC -1- FUNCTION: ALPHA-AMYLASE/TRYPsin INHIBITOR. IT COULD BE INVOLVED
CC IN INSECT DEFENSE MECHANISMS.
CC -1- SUBUNIT: SUBUNIT OF THE TETRAMERIC INHIBITOR.
CC -1- TISSUE SPECIFICITY: ENDOSPERM.
CC -1- MISCELLANEOUS: CM PROTEINS WOULD BE INVOLVED IN THE COOKING
CC QUALITY OF PASTA.
CC -1- SIMILARITY: BELONGS TO THE CEREAL TRYPsin/ALPHA-AMYLASE INHIBITOR
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X17575; CAA35598.1; -
DR PIR: A25310; A25310.
DR PIR: S10027; S10027.
DR HSP: P01087; 181U.
DR InterPro: IPR003612; AAI.
DR InterPro: IPR001768; Cereal_tryp_aml1_inh.
DR Pfam: PF00234; tryp_alpha_aml1. 1.
DR PRINTS: PRO0808; AMYLASEINHTR.
DR SMART: SMO0499; AAI. 1.
DR PROSITE: PS00426; CEREAL_TRYP_AMYL_INH. 1.
KW Serine protease inhibitor; Alpha-amylase inhibitor; Multigene family;
KW Seed; Signal.
FT SIGNAL 1 25
FT CHAIN 26 145 ALPHA-AMYLASE/TRYPsin INHIBITOR CMI.
FT MOD_RS 145 AA; 15517 MW; 119EDBD8C1229672 CRC64;
SQ SEQUENCE 145 AA; 15517 MW; 119EDBD8C1229672 CRC64;

```

```

Query Match 4.3%; Score 7; DB 1; Length 145;
Best Local Similarity 100.0%; Pred. No. 9.1;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 17 LLLATVL 23
DB 10 LLLATVL 16

```

```

RESULT 8
ID URL_ONCMY STANDARD; PRT; 165 AA.
AC 09348;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)

```

```

DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Urotenstin I precursor.
OS Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
CX NCBI_TaxID=8022;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Hypothalamus;
RA Barsyte D., Tipping D., Brennan J., Baker B., Lovejoy D.;
RT "Cloning of rainbow trout urotenstin-I."
RL Submitted (APR-1998) to the EMBL/Genbank/DBJ databases
CC -1- FUNCTION: UROTENSTIN IS FOUND IN THE TELEOST CAUDAL NEUROSECRETORY
CC SYSTEM. IT HAS A SUGGESTED ROLE IN OSMOREGULATION AND AS A
CC CORTICOTROPIN-RELEASING FACTOR. THE NONHORMONAL PORTION OF THIS
CC PRECURSOR MAY BE A UROTENSTIN BINDING PROTEIN, UROPHYSIN.
CC -1- SIMILARITY: BELONGS TO THE SAVIAGINE/CORTICOTROPIN-RELEASING
CC FACTOR/UROTENSTIN I FAMILY OF PEPTIDES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AJ005264; CAA06461.1; -
DR InterPro: IPR00187; CRF.
DR InterPro: IPR003620; Urocortin_CRF.
DR PIR: PF00473; CRF. 1.
DR PRODOM: PD005970; Urocortin_CRF. 1.
DR SMART: SM00039; CRF. 1.
DR PROSITE: PS00511; CRF. 1.
KW Hormone; Cleavage on pair of basic residues; Amidation; Signal.
FT SIGNAL 1 18
FT CHAIN 19 120 UROPHYSIN (POTENTIAL).
FT PEPTIDE 123 163 UROPHYSIN-I.
FT MOD_RS 163 163 AMIDATION (G-164 PROVIDE AMIDE GROUP) (BY
FT SIMILARITY).
SQ SEQUENCE 165 AA; 18631 MW; 3263357EE7653A1A CRC64;

```

```

Query Match 4.3%; Score 7; DB 1; Length 165;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 17 LLLATVL 23
DB 8 LLLATVL 14

```

```

RESULT 9
ID CD3D_RAT STANDARD; PRT; 173 AA.
AC P19377;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE T-cell surface glycoprotein CD3 delta chain precursor (T-cell receptor
DE T3 delta chain).
GN CD3D OR T3D.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
CX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WISTAR;
RX MEDLINE=90356424; Pubmed=2143819;
RA Davies J.D., Mueller D., Wilson D.B., Gold D.P.;
RT "Nucleotide sequence of a cDNA encoding the rat T3 delta chain.";

```

RL Nucleic Acids Res. 18:4617-4617(1990).
 CC -1- FUNCTION: THE CD3 COMPLEX MEDIATES SIGNAL TRANSDUCTION.
 CC -1- SUBUNIT: THE TCR/CD3 COMPLEX OF T LYMPHOCYTES CONSISTS OF EITHER
 CC A TCR ALPHA/BETA OR TCR GAMMA/DELTA HETERODIMER COEXPRESSED AT THE
 CC CELL SURFACE WITH THE INVARIANT SUBUNITS OF CD3 LABELED GAMMA,
 CC DELTA, EPSILON, ZETA, AND ETA.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL: X53430; CAA37521.1; -
 CC DR PIR: S10975; S10975.
 CC DR InterPro: IPR003110; ITAM.
 CC DR Pfam: PF02189; ITAM; 1.
 CC DR SMART: SM00077; ITAM; 1.
 CC KW Immunoglobulin domain; T-cell; Receptor; Transmembrane; Glycoprotein;
 CC Signal.
 CC FT SIGNAL 1 21
 CC FT CHAIN 22 173
 CC FT T-CELL SURFACE GLYCOPROTEIN CD3 DELTA
 CC FT CHAIN.
 CC FT DOMAIN 22 100
 CC FT EXTRACELLULAR (POTENTIAL).
 CC FT TRANSMEM 101 127
 CC FT POTENTIAL.
 CC FT CYTOPLASMIC (POTENTIAL).
 CC FT CARBOHYD 128 173
 CC FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 55 38
 CC FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 74 74
 CC FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC SQ SEQUENCE 173 AA; 19356 MW; C462DFE34E7FDB60 CRC64;

 Query Match 4.3%; Score 7; DB 1; Length 173;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 17 LLLATVYL 23
 DB 10 LLLATVYL 16

 RESULT 10
 ID KAB2.OLDAF STANDARD; PRT; 210 AA.
 AC P58454;
 DT 01-MAR-2002 (Rel. 41, Created)
 DT 01-MAR-2002 (Rel. 41, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Kalata B2 precursor.
 GN OAK4.
 OS Oldenlandia affinis.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; euasterids I; Gentianales; Rubiaceae; Rubioidae;
 OC Spermacoceae; Oldenlandia.
 OC NCBI_TaxID=60225;
 OX [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=21438013; PubMed=11535828;
 RX Jennings C.V., West J., Waine C., Craik D.J., Anderson M.A.;
 RA "Biosynthesis and insecticidal properties of plant cyclotides: the
 RT cyclic knotted proteins from Oldenlandia affinis";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:10614-10619(2001).
 RN [2]
 RP SEQUENCE OF KALATA B2.
 RX MEDLINE=20069551; PubMed=10600388;
 RA Craik D.J., Daly N.L., Bond T., Waine C.;
 RT "Plant cyclotides: a unique family of cyclic and knotted proteins that
 defines the cyclic cystine knot structural motif."

RL J. Mol. Biol. 294:1327-1336(1999).
 CC -1- FUNCTION: Probably participates in a plant defense mechanism. Has
 CC hemolytic activity.
 CC -1- PTM: Kalata B2 is a cyclic peptide.
 CC -1- SIMILARITY: BELONGS TO THE CYCLOTIPE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL: AF393828; AAL05480.1; -
 CC KW Signal; Repeat; Multigene family.
 CC FT SIGNAL 1 22
 CC FT PROPEP 23 66
 CC FT PROPEP 67 95
 CC FT PROPEP 96 120
 CC FT PEPTIDE 121 149
 CC FT PROPEP 150 174
 CC FT PEPTIDE 175 203
 CC FT PROPEP 204 210
 CC FT DISULFID 71 85
 CC FT DISULFID 75 87
 CC FT DISULFID 80 92
 CC FT DISULFID 125 193
 CC FT DISULFID 129 141
 CC FT DISULFID 134 146
 CC FT DISULFID 179 193
 CC FT DISULFID 183 195
 CC FT DISULFID 188 200
 CC SQ SEQUENCE 210 AA; 22327 MW; C6A0D05D294A6147 CRC64;

 Query Match 4.3%; Score 7; DB 1; Length 210;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 28 AFGAEFS 34
 DB 20 AFGAEFS 26

 RESULT 11
 ID YB9K.YEAST STANDARD; PRT; 320 AA.
 AC P38342;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 01-OCT-1994 (Rel. 30, Last annotation update)
 DE Hypothetical 36.0 kDa protein in SHM1-MRP37 intergenic region.
 GN YBR255W OR YBR1734.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OC NCBI_TaxID=4932;
 OX [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=5288C;
 RX MEDLINE=93220397; PubMed=8465606;
 RA Dolignon F., Bileau N., Crouzet M., Aigle M.;
 RT "The complete sequence of a 19,482 bp segment located on the right
 RT arm of chromosome II from Saccharomyces cerevisiae.";
 RL Yeast 9:189-199(1993).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

```

CC  entries requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch)
-----
DR  EMBL; X70529; CAA49930.1; -.
DR  EMBL; Z36134; CAA85228.1; -.
DR  PIR; S32966; S32966.
DR  SGD; S0000469; TSC10.
DR  InterPro: IPR002198; ADH_short.
DR  Pfam; PF00106; adh_short; 1.
KW  Hypothetical protein; Transmembrane.
FT  TRANSMEM 162 182 POTENTIAL.
FT  TRANSMEM 255 275 POTENTIAL.
FT  TRANSMEM 280 300 POTENTIAL.
SQ  SEQUENCE 320 AA; 35986 MW; C5F3E004082FD242 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  110 PKLRGL 116
    |||||
DB  122 PKLRGL 128

RESULT 12
VNS2_AHSV9
ID  VNS2_AHSV9 STANDARD; PRT; 365 AA.
AC  P27279;
DT  01-AUG-1992 (Rel. 23, Created)
DT  01-AUG-1992 (Rel. 23, Last sequence update)
DT  15-DEC-1998 (Rel. 37, Last annotation update)
DE  Nonstructural protein NS2.
GN  S8.
OS  African horse sickness virus 9 (AHSV-9) (African horse sickness virus
OC  (serotype 9)).
OC  Viruses; dsRNA viruses; Reoviridae; Orbiviruses.
OX  NCBI_TaxID=10897;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=92024120; PubMed=1656603;
RA  van Staden V., Theron J., Greyling B.J., Huismans H., Nel L.H.;
RT  "A comparison of the nucleotide sequences of cognate NS2 genes of
RT  three different orbiviruses."
RL  Virology 185:500-504(1991).
CC  -1- FUNCTION: SSRNA-BINDING PROTEIN.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; M69090; AAA42540.1; -.
DR  PIR; A40788; MNXRAH.
KW  Nonstructural protein; RNA-binding.
SQ  SEQUENCE 365 AA; 41193 MW; 45F3A78754887741 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  118 IKYVGS 124
    |||||
DB  121 IKYVGS 127

RESULT 13
RPSD_ANASP
ID  RPSD_ANASP STANDARD; PRT; 390 AA.

```

```

AC  P26683;
DT  01-AUG-1992 (Rel. 23, Created)
DT  01-AUG-1992 (Rel. 23, Last sequence update)
DT  01-MAR-2002 (Rel. 41, Last annotation update)
DE  RNA polymerase sigma factor rpod (Sigma-A).
GN  RPOD OR SIGA OR ALI5263.
OS  Anabaena sp. (strain PCC 7120).
OC  Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX  NCBI_TaxID=103690;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=91193199; PubMed=1901566;
RA  Brahamsa B., Haselkorn R.;
RT  "Isolation and characterization of the gene encoding the principal
RT  sigma factor of the vegetative cell RNA polymerase from the
RT  cyanobacterium Anabaena sp. strain PCC 7120."
RL  J. Bacteriol. 173:2442-2450(1991).
RN  [2]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=21595285; PubMed=11759840;
RA  Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA  Watanabe A., Iriyuchi M., Ishikawa A., Kawashima K., Kimura T.,
RA  Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA  Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
RA  Yasuda M., Tabata S.;
RT  "Complete genomic sequence of the filamentous nitrogen-fixing
RT  cyanobacterium Anabaena sp. strain PCC 7120."
RL  DNA Res. 8:205-213(2001).
CC  -1- FUNCTION: THE SIGMA FACTOR IS AN INITIATION FACTOR THAT PROMOTES
CC  ATTACHMENT OF THE RNA POLYMERASE TO SPECIFIC INITIATION SITES AND
CC  THEN IS RELEASED. THIS IS THE PRIMARY SIGMA-FACTOR OF THIS
CC  BACTERIA.
CC  -----
CC  -1- SIMILARITY: BELONGS TO THE SIGMA-70 FACTOR FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; M60046; AAA22043.1; -.
DR  EMBL; AP003599; BAB76962.1; -.
DR  PIR; A42724; A42724.
DR  HSSP; P00579; 1SIG.
DR  InterPro: IPR000943; Sigma_70.
DR  Pfam; PF00140; sigma70; 1.
DR  PRINTS; PR00046; SIGMA70FCF.
DR  PROSITE; PS00715; SIGMA70_1; 1.
DR  PROSITE; PS00716; SIGMA70_2; 1.
KW  Transcription regulation; Sigma factor; DNA-directed RNA polymerase;
KW  DNA-binding; Complete proteome.
FT  DOMAIN 182 195 POLYMERASE CORE BINDING (POTENTIAL).
FT  DNA_BIND 351 370 H-T-H MOTIF (BY SIMILARITY).
SQ  SEQUENCE 390 AA; 45641 MW; 604814660815284D CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  155 LSEKLR 161
    |||||
DB  120 LSEKLR 126

RESULT 14
DXR_XYLFA
ID  DXR_XYLFA STANDARD; PRT; 396 AA.
AC  Q9PEIO;
DT  16-OCT-2001 (Rel. 40, Created)
DT  16-OCT-2001 (Rel. 40, Last sequence update)

```

01-MAR-2002 (rel. 41, last annotation update)
 1-deoxy-D-xylulose 5-phosphate reductoisomerase (EC 1.1.1.267) (DXP reductoisomerase) (1-deoxyxylulose-5-phosphate reductoisomerase).
 DNR OR XF1048.
 Xylella fastidiosa.
 Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
 Xylella.
 NCBI_TaxId=2371;
 [1]
 SEQUENCE FROM N.A.
 STRAIN=9A5C;
 MEDLINE=20365717; PubMed=10910347;
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M., Alvarenga R., Alves L.M.C., Araya J.E., Bala G.S., Baptista C.S., Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Brites M.R.S., Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carver H., Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M., Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H., Facinani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A., Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R., Gandler M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A., Ho P.L., Holsel J.D., Junqueira M.L., Kemper E.L., Kitaajima J.P., Krieger J.E., Kuramae E.E., Lalget F., Lambais M.R., Leite L.C.C., Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A., Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Matino C.L., Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y., Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B., Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S., Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A., de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A., Peixoto B.R., Pereira G.A.G., Pereira H.A.J.R., Pesquero J.B., Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M., de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E., da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr., da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A., de Souza A.P., Terezi M.F., Truffi D., Tsai S.M., Tsunako M.H., Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L., Zago M.A., Zatz M., Zeldanis J., Setubal J.C.;
 "The genome sequence of the plant pathogen *Xylella fastidiosa*.";
 Nature 406:151-159(2000).
 -1- FUNCTION: CATALYZES THE NADP-DEPENDENT REARRANGEMENT AND REDUCTION OF 1-DEOXY-D-XYLULOSE-5-PHOSPHATE (DXP) TO 2-C-METHYL-D-ERYTHRITOL 4-PHOSPHATE (MEP) (BY SIMILARITY).
 -1- CATALYTIC ACTIVITY: 2-C-methyl-D-erythritol 4-phosphate + NADP(+) = 1-deoxy-D-xylulose 5-phosphate + NADPH.
 -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; second step.
 -1- SIMILARITY: BELONGS TO THE DXR FAMILY.

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

 EMBL: AE003942; AAF3858.1;
 DR InterPro: IPR003821; DXP_reductoisomerase.
 DR Pfam: PF02670; DXP_reductoisom: 1.
 KW Isoprene biosynthesis; Oxidoreductase; NADP; Complete proteome.
 FT NP_BIND 11 18 NADPH (POTENTIAL).
 SQ SEQUENCE 396 AA; 42076 MW; 14B56F95748FD97 CRC64;

Query Match 4.3%; Score 7; DB 1; Length 396;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 14 GKRLLA 20
 |||||
 DB 121 GKRLLA 127

RESULT 15
 ALAT_MUSCR STANDARD; PRT; 412 AA.
 ID ALAT_MUSCR
 AC P26595;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 01-FEB-1994 (Rel. 28, Last annotation update)
 DE Alpha-1-antitrypsin precursor (Alpha-1-antitrypsin) (Alpha-1-proteinase inhibitor) (AAT).
 OS Mus caroli (Wild mouse) (Ricefield mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxId=10089;
 [1]
 SEQUENCE FROM N.A.
 RP MEDLINE=90136592; PubMed=1689000;
 RA Latimer J.J., Berger F.G., Baumann H.;
 RT "Highly conserved upstream regions of the alpha 1-antitrypsin gene in two mouse species govern liver-specific expression by different mechanisms.";
 RT Mol. Cell. Biol. 10:760-769(1990).
 RL -1- FUNCTION: INHIBITOR OF SERINE PROTEASES. ITS PRIMARY TARGET IS ELASTASE, BUT IT ALSO HAS A MODERATE AFFINITY FOR PLASMIN AND THROMBIN.
 CC -1- SUBCELLULAR LOCATION: Extracellular.
 CC -1- TISSUE SPECIFICITY: EXPRESSED NOT ONLY IN LIVER BUT ALSO IN KIDNEY TUBULE CELLS, WHERE IT IS REGULATED BY ANDROGENS DURING DEVELOPMENT.
 CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY.

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

 EMBL: M33567; AAA37128.1;
 DR PIR: A34730; ITMSC.
 DR HSSP: P01009; 9API.
 DR InterPro: IPR000215; Serpin.
 DR Pfam: PF00079; serpin.1.
 DR SMART: SM00093; SERPIN; 1.
 DR PROSITE: PS00284; SERPIN; 1.
 KW Serpin; Serine protease inhibitor; Glycoprotein; Plasma; Signal.
 FT CHAIN 1 24
 FT ACT_SITE 25 412
 FT CARBOHYD 376 377 ALPHA-1-ANTITRYPSINASE.
 FT CARBOHYD 100 100 REACTIVE BOND (BY SIMILARITY).
 FT CARBOHYD 133 133 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 133 133 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 264 264 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 313 313 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 412 AA; 45872 MW; C1A167063CAF4BD CRC64;

Query Match 4.3%; Score 7; DB 1; Length 412;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 20 ATVLQAV 26
 |||||
 DB 368 ATVLQAV 374

Search completed: September 18, 2002, 16:39:30
 Job time: 199 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2002, 16:32:50 ; Search time 16.05 Seconds
(without alignments)
246.539 Million cell updates/sec

Title: US-09-676-718A-1
Perfect score: 162
Sequence: 1 MAAGPSCGLVPAFGKRLILA.....ILKNWTSVEEFLSEKLERI 162

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 231628 segs, 24425594 residues

Word size : 0

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/Backfilist1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7	4.3	346	2	US-08-602-359A-34
2	7	4.3	424	2	US-08-871-268A-23
3	7	4.3	424	3	US-08-871-267B-31
4	7	4.3	424	4	US-09-618-419-31
5	7	4.3	424	4	US-09-163-674-23
6	7	4.3	503	3	US-08-700-651-6
7	7	4.3	510	1	US-08-255-670A-2
8	7	4.3	836	1	US-08-216-971-2
9	7	4.3	836	2	US-08-812-979-2
10	7	4.3	1476	2	US-09-256-703-2
11	7	4.3	1479	2	US-08-951-912-4
12	7	4.3	1479	4	US-09-174-077-4
13	7	4.3	1480	1	US-07-637-621-2
14	7	4.3	1480	1	US-08-136-742A-2
15	7	4.3	1480	1	US-08-135-809A-2
16	7	4.3	1480	1	US-08-466-886-17
17	7	4.3	1480	2	US-08-951-912-2
18	7	4.3	1480	2	US-08-951-912-6
19	7	4.3	1480	2	US-08-469-461-2
20	7	4.3	1480	2	US-08-469-461-4
21	7	4.3	1480	2	US-08-691-605-2
22	7	4.3	1480	2	US-08-455-552A-14
23	7	4.3	1480	3	US-07-890-609-2
24	7	4.3	1480	3	US-07-890-609-4
25	7	4.3	1480	3	US-09-248-026-2
26	7	4.3	1480	4	US-08-469-617-17
27	7	4.3	1480	4	US-08-681-838A-2

28	7	4.3	1480	4	US-08-681-838A-3	Sequence 3, Appl1
29	7	4.3	1480	4	US-09-174-077-2	Sequence 2, Appl1
30	7	4.3	1480	4	US-09-174-077-6	Sequence 6, Appl1
31	7	4.3	1480	5	PCT-US93-11667-2	Sequence 2, Appl1
32	7	4.3	1480	6	5240846-5	Patent No. 5240846
33	6	3.7	24	3	US-08-964-302A-13	Sequence 13, Appl1
34	6	3.7	24	4	US-09-441-416A-13	Sequence 13, Appl1
35	6	3.7	25	1	US-08-366-690-2	Sequence 2, Appl1
36	6	3.7	28	4	US-08-905-223-328	Sequence 328, Appl1
37	6	3.7	50	1	US-08-336-343A-16	Sequence 16, Appl1
38	6	3.7	74	4	US-09-142-565-4	Sequence 4, Appl1
39	6	3.7	95	4	US-09-142-565-6	Sequence 6, Appl1
40	6	3.7	99	3	US-09-173-581-8	Sequence 8, Appl1
41	6	3.7	99	4	US-09-420-915-8	Sequence 8, Appl1
42	6	3.7	131	3	US-09-102-977-2	Sequence 2, Appl1
43	6	3.7	131	4	US-09-034-088A-2	Sequence 2, Appl1
44	6	3.7	132	4	US-09-034-088A-4	Sequence 4, Appl1
45	6	3.7	135	3	US-09-100-391-12	Sequence 12, Appl1

ALIGNMENTS

RESULT 1
US-08-602-359A-34
Sequence 34, Application US/08602359A
Patent No. 5942430
GENERAL INFORMATION:
APPLICANT: ROBERTSON, Daniel E.
APPLICANT: MORPHY, Dennis
APPLICANT: REID, John
APPLICANT: MAFFIA, Anthony
APPLICANT: LINK, Steven
APPLICANT: SMANSON, Ronald V.
APPLICANT: WARREN, Patrick V.
APPLICANT: KOSHOTKA, Anna
TITLE OF INVENTION: ESTERASES
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: FISH & RICHARDSON P.C.
STREET: 4225 EXECUTIVE SQUARE, STE 1400
CITY: LA JOLLA
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/602,359A
FILING DATE: February 16, 1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: HALE, LISA A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 09010/010001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-678-5070
TELEFAX: 619-678-5099
INFORMATION FOR SEQ. ID NO.: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 346 AMINO ACIDS
TYPE: AMINO ACID
TOPOLOGY: LINEAR
MOLECULE TYPE: PROTEIN
US-08-602-359A-34

Query Match 4.3%; Score 7; DB 2; Length 346;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 14 GKRLIA 20
|||
Db 57 GKRLIA 63

RESULT 2

US-08-871-268A-23
; Sequence 23, Application US/08871268A
; Patent No. 5866391
; GENERAL INFORMATION:
; APPLICANT: Jones, Audrey
; APPLICANT: Cherry, Joel R.
; TITLE OF INVENTION: Aspergillus Porphylinogen Synthases
; TITLE OF INVENTION: and Nucleic Acids Encoding Same
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5866391disk of No. 5866391th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/871,268A
; FILING DATE: 09-JUN-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4809,200-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 424 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 5866391e
US-08-871-268A-23

Query Match 4.3%; Score 7; DB 2; Length 424;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 41 LGFSSNL 47
|||
Db 25 LGFSSNL 31

RESULT 3

US-08-871-267B-31
; Sequence 31, Application US/08871267B
; Patent No. 6100057
; GENERAL INFORMATION:
; APPLICANT: Elrod, Susan L.
; APPLICANT: Cherry, Joel R.
; TITLE OF INVENTION: A Method for Increasing Hemoprotein
; TITLE OF INVENTION: Production in Filamentous Fungi
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 6100057o No. 6100057disk of No. 6100057th America, Inc.
; STREET: 405 Lexington Avenue - 64th Fl.
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174

COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/871,267B
; FILING DATE: 9-JUN-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rozek, Carol E.
; REGISTRATION NUMBER: 36,993
; REFERENCE/DOCKET NUMBER: 4771,200-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-878-9652
; TELEFAX: 212-878-9655
; TELEX:
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 424 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 6100057e
US-08-871-267B-31

Query Match 4.3%; Score 7; DB 3; Length 424;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 41 LGFSSNL 47
|||
Db 25 LGFSSNL 31

RESULT 4

US-09-618-419-31
; Sequence 31, Application US/09618419
; Patent No. 6261827
; GENERAL INFORMATION:
; APPLICANT: Elrod, Susan L.
; APPLICANT: Cherry, Joel R.
; TITLE OF INVENTION: A Method for Increasing Hemoprotein
; TITLE OF INVENTION: Production in Filamentous Fungi
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6261827o No. 6261827disk of No. 6261827th America, Inc.
; STREET: 405 Lexington Avenue - 64th Fl.
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/618,419
; FILING DATE: 18-JUL-2000
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/871,267
; FILING DATE: 9-JUN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Rozek, Carol E.

REGISTRATION NUMBER: 36,993
REFERENCE/DOCKET NUMBER: 4771.200-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-878-9652
TELEFAX: 212-878-9655
TELEX: <unknown>
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 424 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: No. 6261827e
SEQUENCE DESCRIPTION: SEQ ID NO: 31:
US-09-618-419-31

Query Match 4.3%; Score 7; DB 4; Length 424;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 41 LGFSSNL 47
|||||||
Db 25 LGFSSNL 31

RESULT 5
US-09-163-674-23
Sequence 23, Application US/09163674
Patent No. 6306630
GENERAL INFORMATION:
APPLICANT: Jones, Aubrey
APPLICANT: Cherry, Joel R.
TITLE OF INVENTION: Aspergillus Porphobilinogen Synthases
TITLE OF INVENTION: and Nucleic Acids Encoding Same
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 63066300 No. 6306630disk of No. 6306630th America, Inc.
STREET: 405 Lexington Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/163,674
FILING DATE: 30-Sept-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Gregg, Valeta A.
REGISTRATION NUMBER: 35,127
REFERENCE/DOCKET NUMBER: 4809.210-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-867-9655
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 424 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: No. 6306630e
US-09-163-674-23

Query Match 4.3%; Score 7; DB 4; Length 424;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 41 LGFSSNL 47
|||||||
Db 25 LGFSSNL 31

RESULT 6
US-08-700-651-6
Sequence 6, Application US/08700651B
Patent No. 6015882
GENERAL INFORMATION:
APPLICANT: PETERSEN, CAROLYN
APPLICANT: LETCH, JAMES
APPLICANT: NELSON, RICHARD, C.
APPLICANT: GUT, JIRI
TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAs AND RNAs
TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
TITLE OF INVENTION: INFECTIONS
FILE REFERENCE: 480.19-4(HV)
CURRENT APPLICATION NUMBER: US/08/700,651B
CURRENT FILING DATE: 1997-08-14
EARLIER APPLICATION NUMBER: 08/415,751
EARLIER FILING DATE: 1995-04-03
NUMBER OF SEQ ID NOS: 15
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 6
LENGTH: 503
TYPE: PRF
ORGANISM: Cryptosporidium parvum
US-08-700-651-6

Query Match 4.3%; Score 7; DB 3; Length 503;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 154 FLESEKLE 160
|||||||
Db 296 FLESEKLE 302

RESULT 7
US-08-255-670A-2
Sequence 2, Application US/08255670A
Patent No. 5691180
GENERAL INFORMATION:
APPLICANT: LOWE, JOHN B
APPLICANT: SMITH, PETER L.
TITLE OF INVENTION: N-ACETYL-GALACTOSAMINE-TRANSFERASE CDNA
TITLE OF INVENTION: SEQUENCE AND EXPRESSION PRODUCTS
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBION, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
STREET: 1755 S. Jefferson Davis Highway, Suite 400
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/255,670A
FILING DATE: 09-JUN-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: OBION, NO. 5691180man F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 2363-092-55
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000

TELEFAX: (703) 413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 510 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-255-670A-2

Query Match 4.3%; Score 7; DB 1; Length 510;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 125 DPVTKLL 131
|||||
DB 55 DPVTKLL 61

RESULT 8
US-08-216-971-2
Sequence 2, Application US/08216971
Patent No. 5639661
GENERAL INFORMATION:
APPLICANT: Welsh, Michael J.
APPLICANT: Sheppard, David N.
TITLE OF INVENTION: NOVEL GENES AND PROTEINS FOR TREATING
TITLE OF INVENTION: CYSTIC FIBROSIS
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street, suite #510
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/216,971
FILING DATE: 23-MAR-1994
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Arnold, Beth E.
REGISTRATION NUMBER: 35,430
REFERENCE/DOCKET NUMBER: UIZ-011
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 836 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-216-971-2

Query Match 4.3%; Score 7; DB 1; Length 836;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 155 LSEKLER 161
|||||
DB 49 LSEKLER 55

RESULT 9
US-08-812-979-2

Sequence 2, Application US/08812979
Patent No. 5958693
GENERAL INFORMATION:
APPLICANT: Welsh, Michael J.
APPLICANT: Sheppard, David N.
TITLE OF INVENTION: NOVEL GENES AND PROTEINS FOR TREATING
TITLE OF INVENTION: CYSTIC FIBROSIS
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street, suite #510
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/812,979
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/216,971
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Arnold, Beth E.
REGISTRATION NUMBER: 35,430
REFERENCE/DOCKET NUMBER: UIZ-011
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 836 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-812-979-2

Query Match 4.3%; Score 7; DB 2; Length 836;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 155 LSEKLER 161
|||||
DB 49 LSEKLER 55

RESULT 10
US-09-256-703-2
Sequence 2, Application US/09256703
Patent No. 6294379
GENERAL INFORMATION:
APPLICANT: Dong, Jian-yun
APPLICANT: Kan, Yuet Wai
TITLE OF INVENTION: The Regents of the University of California
TITLE OF INVENTION: Efficient AAV Vectors
FILE REFERENCE: 023070-08491005
CURRENT APPLICATION NUMBER: US/09/256,703
CURRENT FILING DATE: 1999-02-24
PRIOR APPLICATION NUMBER: US 60/075,980
PRIOR FILING DATE: 1998-02-25
NUMBER OF SEQ ID NOS: 7
SOFTWARE: Patentln Ver. 2.1
SEQ ID NO 2
LENGTH: 1476
TYPE: PRT
ORGANISM: Homo sapiens
OTHER INFORMATION: truncated cystic fibrosis transmembrane

US-09-256-703-2

Query Match 4.3%; Score 7; DB 4; Length 1476;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 155 LSEKLER 161
|||||||
DB 49 LSEKLER 55

RESULT 11

US-08-951-912-4
; Sequence 4, Application US/08951912
; Patent No. 5972995
; GENERAL INFORMATION:
; APPLICANT: Fischer, Horst
; APPLICANT: Illek, Beate
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR CYSTIC
; TITLE OF INVENTION: FIBROSIS THERAPY
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/951,912
; FILING DATE: 16-OCT-1997
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 200116.403
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1479 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-951-912-4

Query Match 4.3%; Score 7; DB 2; Length 1479;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 155 LSEKLER 161
|||||||
DB 49 LSEKLER 55

RESULT 12

US-09-174-077-4
; Sequence 4, Application US/09174077
; Patent No. 6329422
; GENERAL INFORMATION:
; APPLICANT: Fischer, Horst
; APPLICANT: Illek, Beate
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR CYSTIC FIBROSIS THERAPY
; FILE REFERENCE: 200116.403CI
; CURRENT APPLICATION NUMBER: US/09/174,077

; CURRENT FILING DATE: 1998-10-16
; EARLIER APPLICATION NUMBER: US 08/951,912
; EARLIER FILING DATE: 1997-10-16
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1479
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-174-077-4

Query Match 4.3%; Score 7; DB 4; Length 1479;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 155 LSEKLER 161
|||||||
DB 49 LSEKLER 55

RESULT 13

US-07-637-621-2
; Sequence 2, Application US/07637621
; Patent No. 5407796
; GENERAL INFORMATION:
; APPLICANT: cutting, gary
; APPLICANT: antonarakis, stylianos e
; APPLICANT: kazazian jr., haig h
; TITLE OF INVENTION: CYSTIC FIBROSIS MUTATION CLUSTER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie and Beckett
; STREET: 1001 G Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/637,621
; FILING DATE: 19910104
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: kagan, sarah a
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.030010
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9100
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1480 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: HOMO SAPIENS
US-07-637-621-2

Query Match 4.3%; Score 7; DB 1; Length 1480;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 155 LSEKLER 161
|||||||
DB 49 LSEKLER 55

RESULT 14
US-08-136-742A-2
; Sequence 2, Application US/08136742A
; Patent No. 5670488
; GENERAL INFORMATION:
; APPLICANT: Gregory, R.J., Armentano, D., Couture, L.A., Smith,
; TITLE OF INVENTION: GENE THERAPY FOR CYSTIC FIBROSIS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BRUMBAUGH, GRAVES, DONOHUE & RAYMOND
; STREET: 30 ROCKEFELLER PLAZA
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/136,742A
; FILING DATE: 02-DEC-1993
; CLASSIFICATION: 514
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,478
; FILING DATE: 02-DEC-1992
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Seide, Rochelle K.
; REGISTRATION NUMBER: 32,300
; REFERENCE/DOCKET NUMBER: A30668 (Genzyme Dkt. IG4-9.11)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 765-2519
; TELEFAX: (212) 408-2500
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1480 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-136-742A-2

Query Match 4.3%; Score 7; DB 1; Length 1480;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 155 LSEKLER 161
|111111|
DB 49 LSEKLER 55

RESULT 15
US-08-135-809A-2
; Sequence 2, Application US/08135809A
; Patent No. 5688677
; GENERAL INFORMATION:
; APPLICANT: CHENG, SENG H.
; APPLICANT: DITULLIO, PAUL
; APPLICANT: EBBERT, KARL M.
; APPLICANT: MEADE, HARRY M.
; TITLE OF INVENTION: DEOXYRIBONUCLEIC ACIDS CONTAINING
; TITLE OF INVENTION: INACTIVATED HORMONE RESPONSIVE ELEMENTS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENZYME CORPORATION
; STREET: ONE MOUNTAIN ROAD
; CITY: FRAMINGHAM
; STATE: MASSACHUSETTS
; COUNTRY: USA

ZIP: 01701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/135,809A
; FILING DATE: 13-OCT-1993
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: LASSEN, ELIZABETH
; REGISTRATION NUMBER: 31,845
; REFERENCE/DOCKET NUMBER: IG4-9.12
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508) 872-8400
; TELEFAX: (508) 872-5415
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1480 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-135-809A-2

Query Match 4.3%; Score 7; DB 1; Length 1480;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 155 LSEKLER 161
|111111|
DB 49 LSEKLER 55

Search completed: September 18, 2002, 16:37:34
Job time: 284 sec

OM of: US-09-676-718A-1 to: EST: * out_format : pfs

Date: Sep 18, 2002 5:06 PM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-Model=frame+g2n.model -Dev=x1h  
-O=/cgn2.1/USPTO_spool/US09676718/runat.17092002.143432.21293/app-query.fasta.1.222  
-DB=EST -OEMT=fastap -SUFFIX=est -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000 -GAPOP=4.500  
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000  
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELDP=6.000  
-DELXT=7.000 -START=1 -MATRIX=blissum62 -TRANS=human40.cdi  
-LIST=45 -DOCCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0  
-ALIGN=15 -MOD=LOCAL -OUTFMT=pfs -NORM=ext -HEADSIZE=500  
-MINLEN=0 -MAXLEN=200000000 -USER=US09676718 @CON1.1.2515  
-NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30  
-NO_XLPRY -WAIT -THREADS=1
```

Search information block:

Query: US-09-676-718A-1

Query length: 162

Database: EST:*

Database sequences: 13736207

Database length: 1841457050

Search time (sec): 1625.180000

```
score_list:
Sequence      Strd Orig  ZScore  EScore Len  Documentation
gb_est2:BG703377 + 819.00 1699.98 1.9e-85 587 | BG703377 602685165F1 NIH_MGC_95
gb_est2:BG399386 + 819.00 1699.00 2.2e-85 642 | BG399386 602441143F1 NIH_MGC_75
gb_est2:BG779736 + 819.00 1699.63 2.3e-85 664 | BG779736 60268566F1 NIH_MGC_60
gb_est1:AV757896 + 819.00 1698.15 2.4e-85 694 | AV757896 AV757896 BM Homo sapi
gb_est1:AV718270 + 819.00 1698.05 2.5e-85 700 | AV718270 AV718270 FHTB Homo sap
gb_est2:BI518529 + 819.00 1697.46 2.6e-85 726 | BI518529 60306150F1 NIH_MGC_11
gb_est2:BI545979 + 819.00 1697.46 2.7e-85 739 | BI545979 603188102F1 NIH_MGC_95
gb_est2:BG505091 + 819.00 1697.40 2.7e-85 743 | BG505091 60255154F1 NIH_MGC_61
gb_est2:BG613877 + 819.00 1697.38 2.7e-85 744 | BG613877 602639684F1 NIH_MGC_61
gb_est2:BG777365 + 819.00 1697.35 2.7e-85 746 | BG777365 60266422F1 NIH_MGC_61
gb_est2:BI597212 + 819.00 1697.29 2.7e-85 750 | BI597212 603250980F1 NIH_MGC_96
gb_est1:AL533548 + 819.00 1696.77 2.9e-85 787 | AL533548 602549479F1 NIH_MGC_61
gb_est2:BG502649 + 819.00 1696.63 3.0e-85 797 | BG502649 602549479F1 NIH_MGC_61
gb_est2:BG533149 + 819.00 1696.56 3.0e-85 802 | BG533149 60258063F1 NIH_MGC_61
gb_est2:BG529329 + 819.00 1696.48 3.0e-85 808 | BG529329 602585441F1 NIH_MGC_61
gb_est2:BG402494 + 819.00 1696.24 3.1e-85 826 | BG402494 602465503F1 NIH_MGC_73
gb_est2:BG499649 + 819.00 1695.16 3.1e-85 832 | BG499649 602345031F1 NIH_MGC_76
gb_est2:BI597109 + 819.00 1695.85 3.2e-85 853 | BI597109 602351754F1 NIH_MGC_96
gb_est2:BF978924 + 819.00 1695.76 3.3e-85 863 | BF978924 602149392F2 NIH_MGC_62
gb_est2:BE781717 + 819.00 1695.42 3.4e-85 890 | BE781717 601467377F1 NIH_MGC_67
gb_est1:AL559824 + 819.00 1694.15 4.1e-85 999 | AL559824 AL559824 LTI_FLO11_BCI
gb_est2:BM476636 + 819.00 1693.85 4.2e-85 1026 | BM476636 AGENCOURT_6476481 NIH
gb_est2:BG503376 + 816.00 1692.37 5.1e-85 664 | BG503376 602550792F1 NIH_MGC_61
gb_est2:BG612631 + 816.00 1692.27 5.2e-85 669 | BG612631 602640055F1 NIH_MGC_61
gb_est2:BG613959 + 816.00 1690.62 6.4e-85 778 | BG613959 602641880F1 NIH_MGC_61
gb_est2:BG1077461 + 815.00 1689.70 7.2e-85 699 | BG1077461 602664781F1 NIH_MGC_60
gb_est2:BG107890 + 815.00 1688.32 1.4e-84 943 | BG107890 602278158F1 NIH_MGC_84
gb_est2:BG530811 + 813.00 1683.94 1.5e-84 947 | BG530811 602559842F1 NIH_MGC_61
gb_est1:AA305370 + 810.00 1682.25 1.9e-84 531 | AA305370 EST176649 Colon carcin
gb_est1:AA314273 + 810.00 1682.00 2.4e-84 634 | AA314273 EST186117 Colon carcin
gb_est1:AV651776 + 810.00 1680.30 2.4e-84 634 | AV651776 AVE51776 GIC Homo sap
gb_est1:AV707435 + 810.00 1679.85 2.5e-84 661 | AV707435 AV707435 ADB Homo sap
gb_est2:BI668960 + 810.00 1678.22 2.9e-84 719 | BI668960 603296490F1 NIH_MGC_96
gb_est1:AV710046 + 808.00 1676.22 4.0e-84 628 | AV710046 AV710046 Cu Homo sap
gb_est1:BM630212 + 806.00 1674.76 4.9e-84 490 | BM630212 h880a03.y1 NCI CGAP sag
gb_est2:BG714314 + 806.00 1670.03 8.9e-84 754 | BG714314 602669800F1 NIH_MGC_96
gb_est2:BG776825 + 805.00 1668.13 1.1e-83 741 | BG776825 602664005F1 NIH_MGC_59
gb_est2:BF697317 + 799.00 1653.87 7.1e-83 866 | BF697317 602129857F1 NIH_MGC_56
gb_est2:BG433851 + 797.00 1650.09 1.2e-82 1010 | BG433851 AV02497375F1 NIH_MGC_7
gb_est1:AV714563 + 797.00 1654.15 6.9e-83 576 | AV714563 DCB Homo sap
```

```
gb_est2:BG494271 + 796.00 1647.91 1.5e-82 841 | BG494271 602539211F1 NIH_MGC
gb_est2:BG030517 + 793.00 1640.56 3.9e-82 927 | BG030517 602298055F1 NIH_MGC
gb_est2:BG708645 + 792.00 1642.52 3.1e-82 642 | BG708645 602672568F1 NIH_MGC
gb_est2:BG564114 + 792.00 1641.93 3.3e-82 676 | BG564114 602586171F1 NIH_MGC
gb_est2:BI596496 + 792.00 1640.85 3.8e-82 746 | BI596496 60324353F1 NIH_MGC
```

seq_name: gb_est2:BG703377

seq_documentation_block:

```
LOCUS      BG703377          587 bp    mRNA    linear    EST 07-MAY-2001
DEFINITION 602685165F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:4817561 5',
            mRNA sequence.
ACCESSION  BG703377
VERSION    BG703377.1 GI:13975647
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
```

REFERENCE

1 (bases 1 to 587)
NIH-MGC Institute of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.
Email: cgapds-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLAM0718 row: a column: 18

High quality sequence stop: 587.
location/Qualifiers

FEATURES

source

```
1..587
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4817561"
/clone_lib="NIH_MGC_95"
/tissue_type="hippocampus"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescriptR (modified
pbuascript KS+); Site:1: BamHI; Site:2: SalI-XhoI (gtcgag
ct); Oligo-dT primed using primer 5'-tttttttttttttttttttt-3',
size-selected for average insert size 2.5 kb and
normalized to R07 5. This is a primary library enriched
for full-length clones and constructed using the
cap-trapped method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
```

BASE COUNT

156 a 108 c 157 g 166 t

ORIGIN

alignment_scores:

Quality: 819.00 Length: 162

Ratio: 5.119 Gaps: 0

Percent Similarity: 98.765 Percent Identity: 98.765

alignment_block:

US-09-676-718A-1 x BG703377 ..

Align seg 1/1 to: BG703377 from: 1 to: 587

```
1 MetlaalaglyProserGlyCysleuValProalapheglyLysArgLe 17
|||||
30 ATGCGCGCTGGCGCGAGTGTCTGTGCTGCGCGGCTGGCGTACGCTT 79
|||||
17 utleuenaaharhvalleuenaalavalaserialaphethylalagunpes 34
|||||
80 GTTGTGGCGGACTGTGCTTCAAGCGGTGTGCTTGTGGCGGAGGATTTT 129
```

```

34  erSerGluAlaGlySerGluLeuGlyPheSerSerSerLeuLeuGlyCysSer 50
    |||
130  CACGAGAGCATGTCAGAGAGATTAGGCTTTCTAGCAACTGCTTTCAGC 179
    |||
51  SerCysAspLeuLeuGlyInPheAsnLeuLeuGlnLeuAspProAspCys 67
    |||
180  TCTTGTGATCTTCTCGACAGATTCACCTGCTTACAGCTGGATCTGATTG 229
    |||
67  sarGlyGlyCysGlnGluGluAlaGlnPheGlnThrLysLysLeuTyr 84
    |||
230  CAGAGATGCTGTCAGAGAGAAACCAATTTGAACCAAAAAGCTGATG 279
    |||
84  lagLysAlaIleLeuGlnValCysGly***LysLeuGlnYargPheProGln 100
    |||
280  CAGGAGACTATCTTGAAGTTTGATGAAATGGAAAGTTCCCTCA 329
    |||
101  ValGlnAlaPheValArgSerAspLysProLysLeuPheArgLysLeuGln 117
    |||
330  GTCCAGAGCTTTCTGAGAGTGTATAACCAACTGTCAGAGAGACTGCA 379
    |||
117  nileLysTyrValArgLysSerAspProValLeuLysLeuLeuAspAsp 134
    |||
380  AATCAAGTATGTCCGTGTTACAGACCCCTGATTAAGCTTTTGACGACA 429
    |||
134  snGlyAsnIleAlaGlnGluLeuSerIleLeuLysTrpAsnThrAspSer 150
    |||
430  ATGGGAACATGTGCTGAAGACTGAGCATCTTCAATGAACACACAGACAGT 479
    |||
151  ValGlnGluPheLeuSerGluLysLeuGluArgIle 162
    |||
480  GTAGAGAAATCTCTGAGTGAATAAGTTGGAACGACATA 515
    |||
seq_name: gb_est2:BG399386

seq_documentation_block:
LOCUS      BG399386                642 bp      mRNA      linear      EST 12-MAR-2001
DEFINITION 602441143P1 NIH_MGC_75 Homo sapiens CDNA clone IMAGE:4556915 5',
            mRNA sequence.
ACCESSION  BG399386
VERSION    BG399386.1 GI:13292834
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 642)
            NIH-MGC http://mgi.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cga@bbs-remail.nih.gov
            Tissue Procurement: CLOMTECH Laboratories, Inc.
            CDNA Library Preparation: CLOMTECH Laboratories, Inc.
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: L10M1260 row: e column: 12
            High quality sequence stop: 634.
FEATURES
SOURCE
            location/Qualifiers
                1..642
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:4556915"
                /clone_id="NIH_MGC_75"
                /lab_host="DH10B (TI phage-resistant)"
                /note="Organ: kidney; Vector: pDNR-LIB (Clontech); Site_1:
                SfiI (99cgcgcgcgc); Site_2: SfiI (99ccattatggcc); 5' and
                3' adaptors were used in cloning as follows: 5' adaptor
                sequence: 5'-CACGGCATATATGCCC-3' and 3' adaptor sequence:
                5'-ATTCTAGAGCCGAGGCGGCGGACATG-dT(30)BN-3' (where B = A,

```

```

C, or G and N = A, C, G, or T). Average insert size 1.65
kb (range 0.5-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH-MGC Library."
BASE COUNT      166 a      116 c      165 g      195 t
ORIGIN
alignment_scores:
    Quality:      819.00      Length:      162
    Ratio:        5.119      Gaps:        0
    Percent Similarity: 98.765      Percent Identity: 98.765
alignment_block:
US-09-676-718a-1 x BG399386
Align seg 1/1 to: BG399386 from: 1 to: 642
1  MetAlaAlaGlyProSerGlyCysLeuValProAlaPheGlyLysArgIle 17
    |||
15  ATGCGGCGTGGCGCGAGTGGGTGTGTCGTCGCGCGCTTGGGCTACGGTT 64
    |||
17  uleuAlaIthrValLeuGlnAlaValSerAlaPheGlyAlaGlnPheS 34
    |||
65  GTTGTGGCGACGTGCTCTCAACGGGTGCTGCTTTGGGCGACAGCTTTT 114
    |||
34  erSerGluAlaGlyArgGluLeuGlyPheSerSerAsnLeuLeuGlyCysSer 50
    |||
115  CACGAGAGCATGTCAGAGAGATTAGGCTTTCTAGCAACTGCTTTCAGC 164
    |||
51  SerCysAspLeuLeuGlyInPheAsnLeuLeuGlnLeuAspProAspCys 67
    |||
165  TCTTGTGATCTTCTCTCGACAGTTCACCTGCTTTCAGCTGGATCTGATTG 214
    |||
67  sarGlyGlyCysGlnGluGluAlaGlnPheGlnThrLysLysLeuTyr 84
    |||
215  CAGAGATGCTGTCAGAGAGAACCAATTTGAACCAAAAAGCTGATTG 264
    |||
84  lagLysAlaIleLeuGlnValCysGly***LysLeuGlnYargPheProGln 100
    |||
265  CAGGACTATCTTCTGAAGTTTGATGAATAAGTTGGGAAGTCCCTCA 314
    |||
101  ValGlnAlaPheValArgSerAspLysProLysLeuPheArgLysLeuGln 117
    |||
315  GTCCAGAGCTTTGTTGAGTGTATAACCAACTGTCAGAGAGACTGCA 364
    |||
117  nileLysTyrValArgLysSerAspProValLeuLysLeuLeuAspAsp 134
    |||
365  AATCAAGTATGTCCGTGTTACAGACCCCTGATTAAGCTTTTGACGACA 414
    |||
134  snGlyAsnIleAlaGlnGluLeuSerIleLeuLysTrpAsnThrAspSer 150
    |||
415  ATGGGAACATGTGCTGAAGACTGAGCATCTTCAATGAACACACAGACAGT 464
    |||
151  ValGlnGluPheLeuSerGluLysLeuGluArgIle 162
    |||
465  GTAGAGAAATCTCTGAGTGAATAAGTTGGAACGACATA 500
    |||
seq_name: gb_est2:BG779736

seq_documentation_block:
LOCUS      BG779736                664 bp      mRNA      linear      EST 15-MAY-2001
DEFINITION 602668566F1 NIH_MGC_60 Homo sapiens CDNA clone IMAGE:4808032 5',
            mRNA sequence.
ACCESSION  BG779736
VERSION    BG779736.1 GI:14050053
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 664)

```


AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-ri@mail.nih.gov
 Tissue Procurement: DCTD/DRP
 cDNA Library Preparation: CLONETECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: BICM1660 row: d column: 17
 High quality sequence stop: 660.

FEATURES
 source
 1..664
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="4808032"
 /clone_lib="NIH_MGC_60"
 /tissue_type="adenocarcinoma"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: prostate; Vector: pDNR-LIB (Clontech);
 Site.1: SfiI (ggccgctcgcc); Site.2: SfiI (ggccattggcc
 1); Double-stranded cDNA was prepared from cell line RNA.
 5' and 3' adaptors were used in cloning as follows: 5'
 adaptor sequence: 5'-CACGCCATTATGCGC-3' and 3' adaptor
 sequence: 5'-ATCTAGAGCGCGAGCGCGACTG-dT(30)BN-3'
 (where B = A, C, or G and N = A, C, G, or T). Average
 insert size 1.5 kb (range 0.9-4.0 kb). 14/15 colonies
 contained inserts by PCR. This library was enriched for
 full-length clones and was constructed by Clontech
 Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
 library."

BASE COUNT 179 a 118 c 170 g 197 t
ORIGIN

alignment_scores:
 Quality: 819.00 Length: 162
 Ratio: 5.119 Gaps: 0
 Percent Similarity: 98.765 Percent Identity: 98.765

alignment_block:
 US-09-676-718A-1 x BG779736 ..

Align seg 1/1 to: BG779736 from: 1 to: 664

```

1 MetAlaAlaGlyProSerGlyCysLeuValProAlaPheGlyIysArgLe 17
|||||
26 ATGGCGGCTGGCGGAGTGGTGTCTGGTCCGCGCTTGGCGCTACGCTT 75
|||||
17 uleuLeuAlaThrValLeuGlnAlaValSerAlaPheGlyAlaGluPhe 34
|||||
76 GTTGTGGCGACTGTGCTTCAAGCGGTGTCTGCTTTGGGCGAGATT 125
|||||
34 eSerGluAlaCysArgGluLeuGlyPheSerSerAsnLeuLeuCysSer 50
|||||
126 CARGCGAGCATGCAGAGAGTTAGGCTTTCTAGCACTGCTTCGACG 175
|||||
51 SerCysAspLeuLeuGlyGlnPheAsnLeuLeuGlnLeuAspProAsp 67
|||||
176 TCTTGTGATCTTCTCGACAGTTCACCTCTTCAGCTGATCTTGATTG 225
|||||
67 sArgGlyCysCysGlnGluGlnAlaGlnPheGluThrLysLysLeuYrA 84
|||||
226 CAGAGCATGCTGTACAGAGAGACACAAATTGAAACCAAAAGCTGATG 275
|||||
84 laGlyAlaIleLeuGlnValCysGly**LysLeuGlyArgPheProGln 100
|||||
276 CAGGACATATTCTTGAAGTTGTGATGAATAATTGGAGAGTTCCTCAA 325
|||||
101 ValGlnAlaPheValArgSerAspLysProLysLeuPheArgGlyLeuG 117

```

```

|||||
326 GTCACACTTTTGTAGAGATGAACCCCAACTGTTACAGACATCA 375
|||||
117 nIlLysTyValArgLysSerAspProValLeuLysLeuAspAspA 134
|||||
376 AATCAAGATATGTCCTGGTTCAGACCTCTGATTTAAAGCTTTGACGACA 425
|||||
134 snGlyAsnIleAlaGluGluLeuSerIleLeuLysTrpAsnThrAspSer 150
|||||
426 ATGGAGACATTGCTGAAGAACTGAGCACTTCCTCAATGGAACACAGACAGT 475
|||||
151 ValGluGluPheLeuSerGlyLysLeuGlnArgIle 162
|||||
476 GTACAGAAATTCCTGAGTGAAGTGAACGACATCA 511
|||||
seq_name: gb_est1:AV757896

```

seq_documentation_block:

LOCUS AV757896 694 bp mRNA linear EST 19-OCT-2000
DEFINITION AV757896 BM Homo sapiens cDNA clone BMFALD04 5', mRNA sequence.
ACCESSION AV757896
VERSION AV757896.1 GI:10915744
KEYWORDS
 EST.

SOURCE
 ORGANISM

human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
 1 (bases 1 to 694)

AUTHORS
 Gu,Y., Zhao,M., Huang,Q., Xu,X., Li,Y., Peng,Y., Song,H., Xiao,H.,
 Gu,Y., Li,N., Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z., Zeng
 L., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G.,
 Yang,Y., Gao,G., Wang,Z., Zhang,Q., Chen,S., Han,Z. and Chen,Z.

TITLE
 JOURNAL

Unpublished (2000)

COMMENT

Contact: Zenguan Han
 Chinese National Human Genome Center at Shanghai
 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
 201203, P. R. China
 Tel: 86-21-50801919(ex.45)
 Fax: 86-21-50801922
 Email: hanzg@chgc.sh.cn
 This clone is available at CHGC in Shanghai.

FEATURES
 source
 1..694
 Location/Qualifiers

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="BMFALD04"
 /clone_lib="BM"
 /tissue_type="Bone marrow"
 /cell_type="CD34+ hematopoietic stem/progenitor cell"
 /lab_host="BM25.8"
 /note="Vector: pTriplex2; Site.1: sfiI; Site.2: sfiIb"

BASE COUNT 163 a 139 c 189 g 201 t
ORIGIN

alignment_scores:
 Quality: 819.00 Length: 162
 Ratio: 5.119 Gaps: 0
 Percent Similarity: 98.765 Percent Identity: 98.765

alignment_block:
 US-09-676-718A-1 x AV757896 ..

Align seg 1/1 to: AV757896 from: 1 to: 694

```

1 MetAlaAlaGlyProSerGlyCysLeuValProAlaPheGlyIysArgLe 17
|||||
105 ATGGCGGCTGGCGGAGTGGTGTCTGGTCCGCGCTTGGCGCTACGCTT 154
|||||
17 uleuLeuAlaThrValLeuGlnAlaValSerAlaPheGlyAlaGluPhe 34
|||||
155 GTTGTGGCGACTGTGCTTCAAGCGGTGTCTGCTTTGGGCGAGATT 204

```

```

34  eSergluuAaCysArgIuLeuGluPheSerSerAsnLeuLeuCySer 50
|||||
205  CATTGGAGGATGACAGAGATTAGCTTTCTAGCAACTGCTTTCACAC 254
51  SerCysAspLeuLeuGlyInPheAsnLeuGlnLeuAspProAspCy 67
|||||
255  TCTTGATCTTCTCGAGAGTTCAACCTGCTTCAGCTGCATCTGATTG 304
67  sArgGlyCySerGlnGluGlnAlaGlnPheGluThrLysLysLeuTyra 84
|||||
305  CAGAGAGATGCTGTGAGAGGACCAATTGAAACCAAAAGCTGTGANG 354
84  laGlyAlaIleuLeuGluValCysGly**LysLeuGlyArgPheProGln 100
|||||
355  CAGGAGCTATTCTTGAGATTGTGGTGAATAATGGGAAGTTCCCTCAA 404
101  ValGlnAlaPheValArgSerAspLysProLysLeuPheArgGlyLeuGln 117
|||||
405  GTCCAAAGCTTTGTTAGAGAGTGATTAACCAACCAAGCTTCAGAGACTGCA 454
117  nIleLysTyValArgGlySerAspProValLeuLysLeuLeuAspAspA 134
|||||
455  AATCAAGATATGTCGCGTTTCAGACCCCTGTATTAAAGCTTTTGACGACA 504
134  sngLYAsnIleAlaGluGluLeuSerIleLeuLysTyPasnThrAspSer 150
|||||
505  ATGGAACTTCTGTGAGAGACTGAGCATTTCTCAATGGAAACAGACAGACT 554
151  ValGluGluPheLeuSerGluLysLeuGluArgIle 162
|||||
555  GTAGAAGAAATTCCTGAGTGAATAAGTTGGAACGACATA 590

```

seq_name: gb_est1:AV718270

seq_documentation_block:

LOCUS AV718270 700 bp mRNA linear EST 16-OCT-2000
 DEFINITION AV718270 FHTB Homo sapiens cDNA clone FHTBABF07 5', mRNA sequence.
 ACCESSION AV718270
 VERSION AV718270.1 GI:10815422

KEYWORDS EST.
 SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS
 1 (bases 1 to 700)
 Xiao, H., Peng, Y., Song, H., Gu, Y., Yang, Y., Gao, G., Xu, X., Li, N.,
 Qian, B., Liu, F., Qu, J., Gao, X., Cheng, Z., Xu, Z., Zeng, L., Xu, S., Gu
 W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Chen, Z. and Han
 Z.

TITLE
 JOURNAL
 COMMENT
 Homo sapiens cDNA FHTB clones
 Unpublished (2000)

Contact: Zeguang Han

Chinese National Human Genome Center at Shanghai

351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai

201203, P. R. China

Tel: 86-21-50801919 (ex.45)

Fax: 86-21-50801922

Email: hanzg@ncgc.sh.cn

This clone is available at CHGC in Shanghai.

FEATURES

source

1..700

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="FHTBABF07"

/clone_lib="FHTB"

/tissue_type="hypothalamus"

/dev_stage="Fetal"

/lab_host="BM25.8"

/note="Vector: pTribLE2; Site_1: sflIA; Site_2: sflIBI"

BASE COUNT

191 a 130 c 171 g 204 t 4 others

alignment_scores:
 Quality: 819.00 Length: 162
 Ratio: 5.119 Gaps: 0
 Percent Similarity: 98.765 Percent Identity: 98.765

alignment_block:
 US-09-676-718a-1 x AV718270 ..

Align seg 1/1 to: AV718270 from: 1 to: 700

```

1  MetAlaAlaGlyProSerGlyCysLeuValProAlaPheGlyLysArgIle 17
|||||
14  ATGGCGGCTGGGCCGAGCTGGTGTCTGTCGCCGCTTGGCGTACGGTT 63
17  uLeuLeuAlaThrValLeuGlnAlaValSerAlaPheGlyAlaGluPheS 34
|||||
64  GTTGTTGGGAGCTGTGCTTCAAGCGGTGTCTGCTTTGGGCGAGAGTTT 113
34  eSergluuAaCysArgIuLeuGluPheSerSerAsnLeuLeuCySer 50
|||||
114  CATTGGAGGATGACAGAGATTAGCTTTCTAGCAACTGCTTTCACAC 163
51  SerCysAspLeuLeuGlyInPheAsnLeuGlnLeuAspProAspCy 67
|||||
164  TCTTGATCTTCTCGAGAGTTCAACCTGCTTCAGCTGCATCTGATTG 213
67  sArgGlyCySerGlnGluGlnAlaGlnPheGluThrLysLysLeuTyra 84
|||||
214  CAGAGAGATGCTGTGAGAGGACCAATTGAAACCAAAAGCTGTGANG 263
84  laGlyAlaIleuLeuGluValCysGly**LysLeuGlyArgPheProGln 100
|||||
264  CAGGAGCTATTCTTGAGATTGTGGTGAATAATGGGAAGTTCCCTCAA 313
101  ValGlnAlaPheValArgSerAspLysProLysLeuPheArgGlyLeuGln 117
|||||
314  GTCCAAAGCTTTGTTAGAGAGTGATTAACCAACCAAGCTTCAGAGACTGCA 363
117  nIleLysTyValArgGlySerAspProValLeuLysLeuLeuAspAspA 134
|||||
364  AATCAAGATATGTCGCGTTTCAGACCCCTGTATTAAAGCTTTTGACGACA 413
134  sngLYAsnIleAlaGluGluLeuSerIleLeuLysTyPasnThrAspSer 150
|||||
414  ATGGAACTTCTGTGAGAGACTGAGCATTTCTCAATGGAAACAGACAGACT 463
151  ValGluGluPheLeuSerGluLysLeuGluArgIle 162
|||||
464  GTAGAAGAAATTCCTGAGTGAATAAGTTGGAACGACATA 499

```

seq_name: gb_est2:BI518529

seq_documentation_block:

LOCUS BI518529 726 bp mRNA linear EST 29-AUG-2001
 DEFINITION 603061650F1 NIH_MGC_118 Homo sapiens cDNA clone IMAGE:5210821 5',
 mRNA sequence.

ACCESSION BI518529

VERSION BI518529.1 GI:15343321

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 726)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

COMMENT

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.
 Email: cga@ds-remail.nih.gov

Tissue Procurement: Life Technologies, Inc.
 cDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)


```

119 CATGGAGCATGCAGAGAGTAGGTAGGCTTTCTGTGCAACTGCTTTGGAGC 168
116 TCTGTATATCTCTCGACAGTTCAACCTCTTCAGCTGGATCTGTATTG 218
113 TCTGTATATCTCTCGACAGTTCAACCTCTTCAGCTGGATCTGTATTG 218
110 TCTGTATATCTCTCGACAGTTCAACCTCTTCAGCTGGATCTGTATTG 218
107 TCTGTATATCTCTCGACAGTTCAACCTCTTCAGCTGGATCTGTATTG 218
104 TCTGTATATCTCTCGACAGTTCAACCTCTTCAGCTGGATCTGTATTG 218
101 ValGlnAlaPheValArgSerAspIysProIysLeuPheArgIylLeuGI 117
98 GTCCAAACCTTTTGTTAGAGTAGATTAACCCAAACTGTTCAGAGAGCTGCA 368
95 GTCCAAACCTTTTGTTAGAGTAGATTAACCCAAACTGTTCAGAGAGCTGCA 368
92 GTCCAAACCTTTTGTTAGAGTAGATTAACCCAAACTGTTCAGAGAGCTGCA 368
89 GTCCAAACCTTTTGTTAGAGTAGATTAACCCAAACTGTTCAGAGAGCTGCA 368
86 GTCCAAACCTTTTGTTAGAGTAGATTAACCCAAACTGTTCAGAGAGCTGCA 368
83 GTCCAAACCTTTTGTTAGAGTAGATTAACCCAAACTGTTCAGAGAGCTGCA 368
80 GTCCAAACCTTTTGTTAGAGTAGATTAACCCAAACTGTTCAGAGAGCTGCA 368
77 GTCCAAACCTTTTGTTAGAGTAGATTAACCCAAACTGTTCAGAGAGCTGCA 368
74 GTCCAAACCTTTTGTTAGAGTAGATTAACCCAAACTGTTCAGAGAGCTGCA 368
71 GTCCAAACCTTTTGTTAGAGTAGATTAACCCAAACTGTTCAGAGAGCTGCA 368
68 GTCCAAACCTTTTGTTAGAGTAGATTAACCCAAACTGTTCAGAGAGCTGCA 368
65 GTCCAAACCTTTTGTTAGAGTAGATTAACCCAAACTGTTCAGAGAGCTGCA 368
62 GTCCAAACCTTTTGTTAGAGTAGATTAACCCAAACTGTTCAGAGAGCTGCA 368
59 GTCCAAACCTTTTGTTAGAGTAGATTAACCCAAACTGTTCAGAGAGCTGCA 368
56 GTCCAAACCTTTTGTTAGAGTAGATTAACCCAAACTGTTCAGAGAGCTGCA 368
53 GTCCAAACCTTTTGTTAGAGTAGATTAACCCAAACTGTTCAGAGAGCTGCA 368
50 GTCCAAACCTTTTGTTAGAGTAGATTAACCCAAACTGTTCAGAGAGCTGCA 368
47 GTCCAAACCTTTTGTTAGAGTAGATTAACCCAAACTGTTCAGAGAGCTGCA 368
44 GTCCAAACCTTTTGTTAGAGTAGATTAACCCAAACTGTTCAGAGAGCTGCA 368
41 GTCCAAACCTTTTGTTAGAGTAGATTAACCCAAACTGTTCAGAGAGCTGCA 368
38 GTCCAAACCTTTTGTTAGAGTAGATTAACCCAAACTGTTCAGAGAGCTGCA 368
35 GTCCAAACCTTTTGTTAGAGTAGATTAACCCAAACTGTTCAGAGAGCTGCA 368
32 GTCCAAACCTTTTGTTAGAGTAGATTAACCCAAACTGTTCAGAGAGCTGCA 368
29 GTCCAAACCTTTTGTTAGAGTAGATTAACCCAAACTGTTCAGAGAGCTGCA 368
26 GTCCAAACCTTTTGTTAGAGTAGATTAACCCAAACTGTTCAGAGAGCTGCA 368
23 GTCCAAACCTTTTGTTAGAGTAGATTAACCCAAACTGTTCAGAGAGCTGCA 368
20 GTCCAAACCTTTTGTTAGAGTAGATTAACCCAAACTGTTCAGAGAGCTGCA 368
17 GTCCAAACCTTTTGTTAGAGTAGATTAACCCAAACTGTTCAGAGAGCTGCA 368
14 GTCCAAACCTTTTGTTAGAGTAGATTAACCCAAACTGTTCAGAGAGCTGCA 368
11 GTCCAAACCTTTTGTTAGAGTAGATTAACCCAAACTGTTCAGAGAGCTGCA 368
8 GTCCAAACCTTTTGTTAGAGTAGATTAACCCAAACTGTTCAGAGAGCTGCA 368
5 GTCCAAACCTTTTGTTAGAGTAGATTAACCCAAACTGTTCAGAGAGCTGCA 368
2 GTCCAAACCTTTTGTTAGAGTAGATTAACCCAAACTGTTCAGAGAGCTGCA 368
name: gb_est2.BG505091

```

FEATURES
source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4664195"
/clone_lib="NH_MGC_61"
/tissue_type="embryonal carcinoma"
/lab_host="DH10B (T1 phage-resistant)"
/notes="Organ: testis; Vector: PDN-LIB (Clontech); Site:1: SfiI (ggcccttcggccg); Site 2: SfiI (ggcccttcggccg); Double-stranded cDNA was prepared from cell line RNA. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCCGCGAGGGCCGACATC-dT(30)BN-3'

```

(Where B = A', C, or G and N = A, C, G, or T). Average insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library."

BASE COUNT	207	a	133	c	177	g	226	t
ORIGIN								

[illegible]

```
alignment_block:
US-09-676-718A-1 x BG505091
```

Align seg 1/1 to: BG505091 from: 1 to: 743

1 MetAlaIaIaGlyProSerLysIysLeuValProAlaPheGlyLysArgLe 17
15 ATGGCGGGCTGGCCGAGTGGGTGTGTGGCCGGCTTTGGGCTACGGCTT 64
17 uLeuLeuAlaThrValLeuGlnAlaValSerAlaPheGlyIaIaLuePheS 34
65 GTTGTGGCCACGTGCTCTTCACAGCGGTGTCTGCTTTTGGGCAAGATT 114
34 eSerGlnAlaCysArgGluLeuGlyPheSerSerAsnLeuLeuCysSer 50
115 CATCGGAGGCGATCACAAGAGATTAGCTTTTACAGCAACTCTTTCGAGC 164
51 SerCysAspLeuLeuGlyIlnPheAsnLeuLeuGlnLeuLeuAspProAspCy 67
165 TCTGTGATCTTCTGGGCAAGTTCAACCGCTTCACGCTGAGTACCTGATTG 214
67 sArgGlyCysCysGlnGlnGlnAlaGlnPheGlyIuhTrpLysLysLeuIyrA 84
215 CAGAGGATGCTGTCCAGAGAGAACACAAATTTCMAACCAAAAGCTGTATG 264
84 IaGlyAlaIleLeuGlnValaLysGly***LysLeuGlyIArgPheProGln 100
265 CAGAGCATATCTTTCGAAAGCTTTGTGATGAAATAATGGCAAGGTCTCC 314
101 ValGlnAlaPheValaArgSerAspLysProLysLeuPheArgGlyLeuGln 117
315 GTCCAAACCTTTGTGGAGGTATTAACCCAAACGTTCACAGGACAGCA 364
117 nIleLysIyrValaArgGlySerAspProValLeuLysLeuLeuAspAspA 134
365 AATCAACGTATGTCGTGGGTTCACAGACCTGTATTAAGCTTTTGGACACA 414
134 snGlyAsnIleAlaGlnGluLeuSerIleLeuLysIyrPAsnThrAspSer 150
415 ATGGGAACATGCTGTGAAGAACTAGACATCTTCMAATGGAACACAGACAGT 464
151 ValGlnGluPheLeuSerGluLysLeuGlnArgIle 162
165 GTGAGCAATCTCTGTGATAAAAGTTGGACACCA 500

seq_name: gb_est2:BG613877

seq_documentation_block:

LOCUS	BC613877	744 bp	mRNA	linear	EST_18-APR-2001
DEFINITION	602639684.1 NIH_MGC_61 Homo sapiens CDNA clone IMAGE:4770860 5',				
mRNA sequence.					
ACCESSION	BC613877				
VERSION	BC613877.1	GI:13665248			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.				

alignment block:
US-09-676-718a-1 x BG777365 ..

Align seg 1/1 to: BG777365 from: 1 to: 746

```

1 MetAlaAlaGlyProSerGlyCysLeuValProAlaPheGlyLysArgLe 17
  |||
26 ATGGCGGCGTGGCCGAGTGGGTCTGTGGTCCGCGCTTGGGCGTACGGTT 75
  |||
17 uLeuLeuAlaThrValLeuGlnAlaValSerAlaPheGlyAlaGluPheS 34
  |||
76 GTTGTGGGAGCTGTGCTTCAAGCGGTGTCTGCTTTGGGCGAGACTTTT 125
  |||
34 erSerGluAlaCysArgGluLeuGlyPheSerSerAsnLeuLeuGlySer 50
  |||
126 CATCGAGGCGATGCAAGAGATTAGGCTTTCTTACGAACTTGTTCACGC 175
  |||
51 SerCysAspLeuLeuGlyGlnPheAsnLeuGlnLeuAspProGly 67
  |||
176 TCTTGTGATCTTCTCGGACAGTTCAACCTGCTTCAAGCTGATCTGATTG 225
  |||
67 sArgGlyCysCysGlnGluGluAlaGlnPheGluThrLysLysLeuTyr 84
  |||
226 CAGAGGATGCTGTCAAGAGACCAATTGAAACCAAAAGCTGTATG 275
  |||
84 laglYAlaIleLeuGluValCysGly**LysLeuGlyArgPheProGln 100
  |||
276 CAGGAGCTAATCTTGAAGTTGTGATGAATAATTGGGAAGTTCCTCAA 325
  |||
101 ValGlnAlaPheValArgSerAspLysProLysLeuPheArgGlyLeu 117
  |||
326 GTCCAAAGCTTTGTGTAAGAGTATAAACCAACTGTTCAAGAGACTGCA 375
  |||
117 nIleLysTyrValArgGlySerAspProValLeuLysLeuLeuAspAsp 134
  |||
376 AATCAAGATATGTCGGTTCAGACCTGTATTAAGCTTTGGACGACA 425
  |||
134 sngLYAsnIleAlaGluGluLeuSerIleLeuLysTrpAsnThrAspSer 150
  |||
426 ATGGGAACCTTGTGTAAGAGTATTCATAATGGAACACAGACACT 475
  |||
151 ValGluGluPheLeuSerGluLysLeuGluArgIle 162
  |||
476 GTAGAAGAAATTCCTGAGTCAAAAGTTGGAACGACATA 511
  |||
seq_name: gb_est2:BI597212

seq_documentation_block:
LOCUS      BI597212              750 bp      mRNA      linear      EST 07-SEP-2001
DEFINITION 603250980F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:5302891 5',
            mRNA sequence.
ACCESSION  BI597212
VERSION    BI597212.1  GI:15490151
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 750)
            NIH-MGC http://mgs.ncl.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: c9apds-remail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NMGRI), Shiraki
            Toshivuki and Piero Carninci (RIKEN)
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LNL at:
            http://image.lnl.gov
            Plate: LHAM1766  row: 0  column: 20

```

High quality sequence stop: 747.
Location/Qualifiers
1..750
source

```

FEATURES
    source
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="IMAGE:5302891"
        /clone_id="NIH_MGC_96"
        /tissue_type="hypothalamus"
        /lab_host="DH10B"
        /note="Organ: brain; Vector: pBluescriptR (modified
        pBluescript KS+); Site.1: BamHI; Site.2: SalI-XhoI (gcgag
        ); Oligo dt primed using primer 5'-TTTTTATTTTATTTT-3',
        size-selected for average insert size 2.3 kb and
        normalized to R0.5. This is a primary library enriched
        for full-length clones and constructed using the
        Cap-trapper method (Carninci, in preparation). Library
        constructed by M. Brownstein (NIH/NMGRI, National
        Institutes of Health). Note: this is a NIH-MGC Library."
BASE COUNT      202 a      140 c      183 g      225 t
ORIGIN

```

alignment_scores:
Quality: 819.00 Length: 162
Ratio: 5.119 Gaps: 0
Percent Similarity: 98.765 Percent Identity: 98.765

alignment block:
US-09-676-718a-1 x BI597212 ..

Align seg 1/1 to: BI597212 from: 1 to: 750

```

1 MetAlaAlaGlyProSerGlyCysLeuValProAlaPheGlyLysArgLe 17
  |||
34 ATGGCGGCGTGGCCGAGTGGGTCTGTGGTCCGCGCTTGGGCGTACGGTT 83
  |||
17 uLeuLeuAlaThrValLeuGlnAlaValSerAlaPheGlyAlaGluPheS 34
  |||
84 GTTGTGGGAGCTGTGCTTCAAGCGGTGTCTGCTTTGGGCGAGACTTTT 133
  |||
34 erSerGluAlaCysArgGluLeuGlyPheSerSerAsnLeuLeuGlySer 50
  |||
134 CATCGAGGCGATGCAAGAGATTAGGCTTTGTGTAAGACTTGTTCACGC 183
  |||
51 SerCysAspLeuLeuGlyGlnPheAsnLeuGlnLeuAspProGly 67
  |||
184 TCTTGTGATCTTCTCGGACAGTTCAACCTGCTTCAAGCTGATCTGATTG 233
  |||
67 sArgGlyCysCysGlnGluGluAlaGlnPheGluThrLysLysLeuTyr 84
  |||
234 CAGAGGATGCTGTCAAGAGACCAATTGAAACCAAAAGCTGTATG 283
  |||
84 laglYAlaIleLeuGluValCysGly**LysLeuGlyArgPheProGln 100
  |||
284 CAGGAGCTAATCTTGAAGTTGTGATGAATAATTGGGAAGTTCCTCAA 333
  |||
101 ValGlnAlaPheValArgSerAspLysProLysLeuPheArgGlyLeu 117
  |||
334 GTCCAAAGCTTTGTGTAAGAGTATTAACCAACTGTTCAAGAGACTGCA 383
  |||
117 nIleLysTyrValArgGlySerAspProValLeuLysLeuLeuAspAsp 134
  |||
384 AATCAAGATATGTCGGTTCAGACCTGTATTAAGCTTTGGACGACA 433
  |||
134 sngLYAsnIleAlaGluGluLeuSerIleLeuLysTrpAsnThrAspSer 150
  |||
434 ATGGGAACCTTGTGTAAGAGTATTCATAATGGAACACAGACACT 483
  |||
151 ValGluGluPheLeuSerGluLysLeuGluArgIle 162
  |||
484 GTAGAAGAAATTCCTGAGTCAAAAGTTGGAACGACATA 519
  |||
seq_name: gb_est1:AL5353548

```

```

seq_documentation_block:
LOCUS      AL533548              787 bp      mRNA      linear      EST 13-FEB-2001
DEFINITION AL533548 LTI_FL015_Brn1 Homo sapiens cDNA clone CS0DN004YB08 5
prime, mRNA sequence.
ACCESSION  AL533548
VERSION     AL533548.1  GI:12797041
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 787)
AUTHORS     Li, W.-B., Gruber, C., Jesse, J. and Polayes, D.
TITLE       Full-length cDNA libraries and normalization
JOURNAL     Unpublished (2001)
COMMENT     Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, web : www.genoscope.cns.fr.

FEATURES
source
1..787
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="LTI_FL015_Brn1"
/sex="male"
/note="Vector: pCMVSPORT 6; site_1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was constructed by Life Technologies.
Contact : Feng Liang Life Technologies, a division of
Invitrogen 9800 Medical Center Drive Rockville, Maryland
20850, USA Fax : (1) 301 610 8371 Email :
fliang@litech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT      216 a      140 c      185 g      246 t

alignment_scores:
Quality:      819.00      Length:      162
Ratio:        5.119      Gaps:      0
Percent Similarity: 98.765      Percent Identity: 98.765

alignment_block:
US-09-676-718a-1 x AL533548 ..

Align seg 1/1 to: AL533548 from: 1 to: 787
1 MetAlaAlaGlyProSerGlyCysLeuValProAlaPheGlyLysArgIle 17
|||||
6 ATGGCGGCTGGGCGGAGTGGGTCTGTGGTCCGGCGCTTTGGGCTACGGTT 55
|||||
17 uleuLeuAlaIhrValLeuGlnAlaValSerAlaPheGlyAlaGluPheS 34
|||||
56 GTTGTGGGAGCTGTGCTTCAAGCGGTGCTGCTTTGGGGCAGAGTTT 105
|||||
34 eSerGlnAlaGlySarGluLeuGlyPheSerSerAsnLeuLeuCysSer 50
|||||
106 CATCGAGGACATGCAGAGAGTAGGCTTTTCAGCAACTGCTTTCACAG 155
|||||
51 SerCysAspLeuLeuGlyGlnPheAsnLeuGlnLeuAspProAspCy 67
|||||
156 TCTTGTGATCTTCTCGGACAGTTCAACCTGCTTCACTGATCTCATTTG 205
|||||
67 sArgGlyCysGlnGluGluAlaGlnPheGluIhrLysLysLeuTyrA 84
|||||
206 CAGAGATGCTGTCAGAGAGAACAAATTTGAAACCAAAAAGCTGATG 255
|||||
84 lAgLyAlaIleuGlnValGlyGly**LysLeuGlyArgPheProGln 100

```

```

seq_documentation_block:
LOCUS      BG502649              797 bp      mRNA      linear      EST 27-MAR-2001
DEFINITION 602549479P1 NIH_MGC_61 Homo sapiens cDNA clone IMAGE:4657338 5',
mRNA sequence.
ACCESSION  BG502649
VERSION     BG502649.1  GI:13464166
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 797)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
Email: cga@bs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: lMCM1449 row: m column: 19
High quality sequence stop: 753.

FEATURES
source
1..797
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="IMAGE:4657338"
/clone_id="NIH_MGC_61"
/tissue_type="embryonal carcinoma"
/lab_host="DH10B (TI Phage-resistant)"
/note="Organ: testis; Vector: pDNR-LIB (Clontech); Site_1:
SfiI (ggccgcctcgcc); Site_2: SfiI (ggccatcattggcc);
Double-stranded cDNA was prepared from cell line RNA.
5' and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-ATTCTAGAGCGCCGAGCGCCGACATG-dT(30)BN-3'
(where B = A, C, or G and N = A, C, G, or T). Average
insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
library."

BASE COUNT      218 a      141 c      192 g      246 t

ORIGIN
1
|||||
alignment_scores:
Quality:      819.00      Length:      162

```


151 ValGluGluPheLeuSerGluLysLeuGluArgIle 162
 |||||||
 484 GTAGAGAATTCTCGAGTGAATAAGTGAACGCATA 519

seq_name: gb_est2:BG529329

seq_documentation_block:

LOCUS BG529329 808 bp mRNA linear EST 03-APR-2001
 DEFINITION 60258441F1 NIH_MGC_61 Homo sapiens cDNA clone IMAGE:4687019 5',
 mRNA sequence.

ACCESSION BG529329

VERSION BG529329.1 GI:13520866

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 808)

AUTHORS NIH-MGC http://mgs.ncl.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs-r@mail.nih.gov
 Tissue Procurement: ATCC

cDNA Library Preparation: CLONTECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov

Plate: LICM1499 row: b column: 12
 High quality sequence stop: 732.
 Location/Qualifiers

FEATURES
 source

1..808
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="4687019"
 /clone_lib="NIH_MGC_61"
 /tissue_type="embryonal carcinoma"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: testis; Vector: pDNR-LIB (Clontech); Site_1:
 SfiI (ggcgcttggc); Site_2: SfiI (ggcgcttggc);
 double-stranded cDNA was prepared from cell line RNA. 5'
 and 3' adaptors were used in cloning as follows: 5'
 adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor
 sequence: 5'-ATTCTAGAGCGCGAGCGGCACATG-dT(30)BN-3'
 (where B = A, C, or G and N = A, C, G, or T). Average
 insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies
 contained inserts by PCR. This library was enriched for
 full-length clones and was constructed by Clontech
 Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
 library."

BASE COUNT 222 a 145 c 200 g 241 t
 ORIGIN

alignment_scores:

Quality: 819.00 Length: 162
 Ratio: 5.119 Gaps: 0
 Percent Similarity: 98.765 Percent Identity: 98.765

alignment_block:

US-09-676-718a-1 x BG529329 ..

Align seg 1/1 to: BG529329 from: 1 to: 808

1 MetAlaAlaIyProSerGlyCysLeuValProAlaPheGlyLysArgIle 17
 |||||||
 16 ATGGCGCGCTGGCGCGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 65
 |||||||
 17 uLeuLeuAlaThrValLeuGlnAlaValSerAlaPheGlyValAlaPheS 34
 |||||||
 66 GTTGTGGCAGACTGTCTTCAAGCGGTGTCTTTTGGGCGAGAGTTT 115

34 eRSerGluAlaCysArgGluLeuGlyPheSerSerAsnLeuLeuCysSer 50
 |||||||
 116 CATCGAGAGCATCGACAGAGATTAGCTTTTCTACCAACTCTCTTGCACG 165
 |||||||
 51 SerCysAspLeuLeuGlyGlnPheAsnLeuLeuGlnLeuAspProAspC 67
 |||||||
 166 TCTTGTGATCTTCTCGGACAGTTCACACTGCTTCAAGTGTGATGATG 215
 |||||||
 67 sArgGlyCysCysGlnGluGluAlaGlnPheGluThrLysLysLeuTyrA 84
 |||||||
 216 CAGAGATGCTGTGAGAGAGACACAAATTGAAACCAAAAAGCTGTATG 265
 |||||||
 84 laGlyAlaIleLeuGluValCysGly**LysLeuGlyArgPheProGln 100
 |||||||
 266 CAGAGCTATTCTTGAAGTTTGTGATGAATAATTGGCAAGTTCCTCCCA 315
 |||||||
 101 ValGlnAlaPheValArgSerAspLysProLysLeuPheArgGlyLeuG 117
 |||||||
 316 GTCCAGCTTTTGTAGAGATGAATTAACCAACTGTTCAAGAGACTGCA 365
 |||||||
 117 nileLysTyrValArgGlySerAspProValLeuLysLeuLeuAspAspA 134
 |||||||
 366 AATCAAGTATGTCCGTGTGAGACCTGTATTAAGCTTTTGGACGACA 415
 |||||||
 134 snGlyAsnIleAlaGluGluLeuSerIleLeuLysTrpAsnThrAspSer 150
 |||||||
 416 ATGGGAACATGTGTGAAGAACTGAGCAATTTCAAATGGAACACAGACAGT 465
 |||||||
 151 ValGluGluPheLeuSerGluLysLeuGluArgIle 162
 |||||||
 466 GTAGAGAATTCTCGAGTGAATAAGTGAACGCATA 501

THIS PAGE BLANK (USPTO)


```

343 GTCACANCTTTGTGTANGATGATATAAACCAACTGTTACAGAGCTG 392
116 euglnlilelystyryal.ArgglyseraspprovalleuLysleuLeu 132
333 CCAATTAATAAGTATGTNNCGTGTTCACAAACGTCGATTAATAAGCTTT 442
132 PASP 133
443 NGAC 446

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-2

seq_documentation_block:
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
US-09-103-840A-2

alignment_scores:
Quality: 78.00 Length: 141
Ratio: 1.000 Gaps: 6
Percent Similarity: 55.319 Percent Identity: 26.950

alignment_block:
US-09-676-718A-1 x US-09-103-840A-2 ..

Align seg 1/1 to: US-09-103-840A-2 from: 1 to: 4403765

12 Alapheglylvsargleu.....leu 20
503840 ACCTTCGCTCCGCGCTCATCTAGGTACTGGGGTGCAGCAATCTGTC 503889
20 athrvalleuglnalavalseralaphelylagluhpheSerSerGlu 37
503890 GGTCTAGAGGAGCGCTGATGCGCTCAGGTACGAGCTGACACCGCTG 503939
37 lacysargluleucllyphe.....SerSeransleucllySerSer 51
503940 CGATACGCGCGGCTCGACCGGAGGAGGACCGGCTGCTC..... 503980
52 CysaspleuLeuGlnPheansleuLeuGlnleuAspprovalleu 68
503981 ..GACCTGCTACCGGCTCGGCTACACCGCTA...CCCAACACCGC 504024
68 ggllycscysglnlgluLalaglnPhegluthrlyslsleuTYAlag 85
504025 GGGGTCCGCGACGCGCGGGAAGCGGTCTGACAGCGAGTGGCGCCGTC 504074
85 lyAlaile.....leuGlnValcysgly**tysleu 95
504075 AGGCGCTGAACCAACTGGGTCAAGCTCGAGGTGATTCGCGACGAAAC 504124
96 GlyArgPheProGlnValGlnAlaPheValArgSerAspLysProLysle 112

```

```

504125 ACCGTGTGGCTGATCCGTCGCAATTAGTCGCGGTGCAGAACAAATTG 504174
112 upheargglyleuGlnlilelystyryalArgglyseraspprovalleu 128
504175 GGACGACGATTTGTCCTTACCGTACGACAAACGACGACCGGCTCTG 504224
129 ..lysleuLeuAspAspansgly 135
504225 CCCGCCGCTAGAGATACCGGT 504247

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-972-927-1

seq_documentation_block:
; Sequence 1, Application US/08972927
; Patent No. 6166290
; GENERAL INFORMATION:
; APPLICANT: Rea, Philip A
; APPLICANT: Lu, Yu-Ping
; APPLICANT: Li, Ze-Sheng
; TITLE OF INVENTION: GLUTATHIONE-S-CONJUGATE TRANSPORT IN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESS: PANITCH SCHWARZE JACOBS & NADEL, P.C.
; STREET: One Commerce Square, 2005 Market Street, 22nd
; STREET: Floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: US
; ZIP: 19103-7086
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/972,927
; FILING DATE: 18-NOV-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/031,040
; FILING DATE: 18-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/061,328
; FILING DATE: 08-OCT-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Doyle Leary Ph.D., Kathryn
; REGISTRATION NUMBER: 36,317
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-565-1284
; TELEFAX: 215-567-2991
; TELEX: 831-494
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5232 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-972-927-1

alignment_scores:
Quality: 77.50 Length: 85
Ratio: 1.615 Gaps: 2
Percent Similarity: 56.471 Percent Identity: 28.235

alignment_block:
US-09-676-718A-1 x US-08-972-927-1/rev ..

Align seg 1/1 to reverse of: US-08-972-927-1 from: 1 to: 5232

```

```

32 GIupheSerSerGluAlaCysArgGluLeuGlyPheSerSerAsnLeu 48
   ||||| ||||| : : ||||| : ||||| :
4414 GAATCTTCTCGATAGTCTCTGAAATGAGGCGCATCGTTCTAACATCTA 4365
   ||||| ||||| : : ||||| : ||||| :
48 ucysSerSerCysAspLeuLeuGlyGlnPhe..... 58
   ||||| ||||| : : ||||| : ||||| :
4364 CAGCAGCACTGCTTCATCAGAGCAGATCTTAGATCTCCGTAACAC 4315
   .....AsnLeuLeuGlnLeuAspProAspCysArgGlyCys 70
   : : ||||| : : ||||| :
4314 GCACGTGAAGACTCAACAATTGCTCTCTCCACGCTGAATTCCTCC 4265
   71 CysGlnGluGlnAlaGlnPheGlnThrLysLysLeuTrpAlaGlyAla 87
   || : : : : : : ||||| : : ||||| :
4264 TCCTCAGACCTCAGCATCAAGACCAAGAGATTTCTCGGATGCTAT 4215
   87 eleuGluValCysGly**LysLeuGlyLysPheProGlnValGlnAla 104
   ||||| : ||||| ||||| ||||| : : : :
4214 CTTTCAAGTGTG.....CTCTCTAGAGATTCCTCAAGATCAGCATCA 4171
   104 heval 105
   : : |||
4170 TTGTG 4166

```

seq_name: /cgn2_6/ptodata/2/1na/5B_COMB.seq:US-08-912-227-1

```

seq_documentation_block:
; Sequence 1, Application US/08912227
; Patent No. 5998171
; GENERAL INFORMATION:
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Ni, Jian
; APPLICANT: Rosen, Craig A.
; TITLE OF INVENTION: Human Endokine Alpha
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P. L. L. C
; STREET: 1100 NEW YORK AVE., NW, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/912,227
; FILING DATE: Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/024,058
; FILING DATE: 16-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Stefie, Eric K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488, 0470001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2340
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1849 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 53...559
; US-08-912-227-1

```

```

alignment_scores:
  Quality: 75.00      Length: 98
  Ratio: 1.562
  Percent Similarity: 48.980  Percent Identity: 32.653

```

alignment_block:
US-09-676-718a-1 x US-08-912-227-1 ..

Align seg 1/1 to: US-08-912-227-1 from: 1 to: 1849

```

5 ProSerGlyCysLeuValProAlaPheGlyLysArgLeuLeuAla.. 20
   ||| ||| ||| ||| : : ||||| :
1536 CCCCTCATTGCACTGCTCAGCCTTAGAATGTATTCTCTATCGTC 1585
   21 .....ThrValLeuGlnAlaValSera 28
   ||||| : ||||| :
1586 TTATTAAGACTTAAACTTAGCAATTCATTTTCAGGAAGCATATTCC 1635
   28 laPheGlyAlaGluPheSerSerGluAlaCysArgGluLeuGlyPheSer 44
   : : ||||| : : ||||| :
1636 CTTTACCCAGGTGAGCAGAGTGAGCTACACAGATCTTCTTACC 1685
   45 SerAsnLeuLeuCysSerSerCysAspLeuLeuGlyGlnPheAsnLeu 61
   ||||| : ||||| :
1686 AGCACACTTTTCTTTTCTCTC.....CTGAA 1714
   61 uGlnLeuAspProAspCysArgGlyCysGlnGluGlnAlaGlnPhe 77
   ||| ||||| ||||| ||||| :
1715 TCAGGAGATCCA.....GGATGCTGTTCAGGCGCTTATCCAAACA 1755
   78 GluThrLysLysLeuTrpAlaGlnAlaIleLeuGlnValCys 91
   : : ||||| : ||||| :
1756 AATTCCCTCTTCATCTTGAGGCGCCCATTTAGTCAAAAGT 1797
   seq_name: /cgn2_6/ptodata/2/1na/6B_COMB.seq:US-08-972-927-4

seq_documentation_block:
; Sequence 4, Application US/08972927
; Patent No. 6166290
; GENERAL INFORMATION:
; APPLICANT: Rea, Philip A
; APPLICANT: Lu, Yu-Ping
; APPLICANT: Li, Ze-Sheng
; TITLE OF INVENTION: GLUTATHIONE-S-CONJUGATE TRANSPORT IN
; TITLE OF INVENTION: PLANTS
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P. C.
; STREET: One Commerce Square, 2005 Market Street, 22nd
; STREET: Floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: US
; ZIP: 19103-7086
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/972,927
; FILING DATE: 18-NOV-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/031,040
; FILING DATE: 18-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/061,328
; FILING DATE: 08-OCT-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Doyle Leary Ph.D., Kathryn

```

```
REGISTRATION NUMBER: 36,317
REFERENCE/DOCKET NUMBER: 9596-1202
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-965-1284
TELEFAX: 215-567-2991
TELEX: 831-494
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 5175 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-972-927-4

alignment_scores:
Quality: 74.50 Length: 85
Ratio: 1.585 Gaps: 2
Percent Similarity: 55.294 Percent Identity: 25.882

alignment_block:
US-09-676-718A-1 x US-08-972-927-4/rev ..

Align seg 1/1 to reverse of: US-08-972-927-4 from: 1 to: 5175

32 GluPheSerSerGluAlaCysArgGluLeuGlyPheSerSerAsnLeuLe 48
||||| ||||| : : : : : ||||| : :
4385 GAATTCCTCTCGAATGGTCTTGGATGAGAACATCACTTCTTACGTCA 4336
||||| ||||| : : : : : ||||| : :
48 uCysSerSerCysAspLeuLeuGlyGlnPheAsnLeuLeuAsp 65
||||| ||||| : : : : : ||||| : :
4335 CTGCAGCAGTGTCTTCACAGAACAGATATCTAGATCTCGTAACAT 4286
65 roAspCysArgGly.....Cys 70
|||||
4285 GCAGCTGCACACATCAACACTGCTCTGCCAACACGAAATCTCTCC 4236
71 CysGlnGluGlnAlaGlnPheGluThrLysLysLeuTyralaGlyAla 87
||||| : : : : : ||||| : :
4235 TGCCTTGATCTACCTCAGCATCAGCAGCAAGATTTCTCGCGATAGTAT 4186
87 LeuGlnValCysGly**LysLeuGlyArgPheProGlnValGlnAla 104
| : : : : | ||||| : : : :
4185 CTTCAGATGTGCCCTCTCAG.....AGATCCACAGAGATCGCGCTGC 4142
104 heVal 105
|||
4141 TTGTG 4137

seq_name: /cgn2_6/plodata/2/1na/5A_COMB.seq:US-08-441-139-6

seq_documentation_block:
; Sequence 6, Application US/08441139
; Patent No. 5773245
; GENERAL INFORMATION:
; APPLICANT: Wiltrop, Dr. Karl D.
; APPLICANT: Robinson, Anne S.
; TITLE OF INVENTION: METHODS FOR INCREASING SECRETION OF
; TITLE OF INVENTION: RECOMBINANTLY EXPRESSED PROTEINS
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCUDLY, SCOTT, MURPHY & PRESSER
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: NY
; COUNTRY: USA
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
```

```
CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/441,139
; FILING DATE: 15-MAY-1995
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/089,997
; FILING DATE: 06-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 8646
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 516-742-4343
; TELEFAX: 516-742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2574 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 441..2429
; US-08-441-139-6

alignment_scores:
Quality: 74.00 Length: 94
Ratio: 1.396 Gaps: 3
Percent Similarity: 56.383 Percent Identity: 27.660

alignment_block:
US-09-676-718A-1 x US-08-441-139-6 ..

Align seg 1/1 to: US-08-441-139-6 from: 1 to: 2574

53 AspleuLeuGlyGlnPheAsnLeuLeuAspProAspCysArgG1 69
||||| : : : : : |||||
1881 AACCTTGTGGAATATTGACCTTGATGTAATCTCTGCCCTCGTGG 1930
69 yCysCysGlnGluAlaGlnPheGluThrLysLysLeuTyralaGly 86
| : : : : | ||||| : :
1931 TGTTCGCCCAATTCAGATGACGTTGAAGTCGAT.....GCCAATG 1971
86 laileuGluValCysGly...LysLeuGlyArgPheProGlnVal 101
| : : : : | ||||| : :
1972 GTGTTTACTGTTTCAGCCGTCGACCAAGCTCTGTAAGGTAAGCTGAG 2021
102 GlnAlaPheValArgSerAspLysProLysLeuPheArgGlyLeuGln 118
| : : : : | ||||| : :
2022 AAGCTTGTATCAGATGACCAAGGTCGTTG..... 2054
118 eluSTyValArgGlySerAspProValLeuLysLeuAspAspAng 135
| : : : : | ||||| : :
2055 .....TCTGAGGAAGATATCGAGCGCATGCTTAAGAGCGCG 2091
135 LysAlaLeuGluGluSerLeuLys 145
| : : : : | ||||| : :
2092 AAGATTCGCTGAGAGATAGATTTTGAAG 2123

seq_name: /cgn2_6/plodata/2/1na/5A_COMB.seq:US-08-178-477B-42

seq_documentation_block:
; Sequence 42, Application US/08178477B
; Patent No. 5756343
; GENERAL INFORMATION:
; APPLICANT: WU, CARL, CLOS, JOACHIM,
; APPLICANT: WESTWOOD, J. TIMOTHY.; RABINDRAN, SRIDHAR
; TITLE OF INVENTION: CELL STRESS
; TITLE OF INVENTION: TRANSCRIPTIONAL FACTORS
; NUMBER OF SEQUENCES: 42
```

```

CORRESPONDENCE ADDRESS:
ADDRESSER: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154

COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/178,477B
FILING DATE: 07-JAN-1994
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/617,910
FILING DATE: 26-NOV-1990
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: CAROL M. GRUPPI
REGISTRATION NUMBER: 37,341
REFERENCE/DOCKET NUMBER: 2026-4103US1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 751-6849
TELEFAX: (212) 751-4800
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 2781
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-178-477B-42

alignment_scores:
Quality: 69.50      Length: 167
Ratio: 0.827        Gaps: 9
Percent Similarity: 50.299   Percent Identity: 27.545

alignment_block:
US-09-676-718a-1 x US-08-178-477B-42/rev ..
Align seg 1/1 to reverse of: US-08-178-477B-42 from: 1 to: 2781

13 PhgGLyLysArgLeuLeuAlaThrValLeuGlnAlaValSer.... 27
1919 TTCTCAGCCGCACTGTATTGTCGACAGCCATCCCTCTCCACAGATGC 1870
28 ....AlaPhgGLyAlaGluPheSerSerGlnAlaCysArgGluLeuGlyP 43
1869 TTCCGCGCTTGCCGCTGCTTCATGTCGAGGCC.....TGATTGGCAT 1826
43 heSerSerAsnLeuLeuCysSerSerCysAspLeuLeuGlnPheAsn 59
1825 CCAGCAGCGACTGTTGGTGGCAGCATCGAAGTGTGCGCA.....AAC 1782
60 LeuLeuGlnLeuAspProAspCysAlaGly..CysCysGlnGluGlu.. 74
1781 TTGCCCCCTGCTCACTGTCGATTCGATCCCAATTTTGTGCTGCTGTCAGTTT 1732
75 ....AlaGlnPheGluThrLysLysLeuTyraGlnAlaIleLeuGluV 90
1731 GGTGCGCTTTCACAGCCGCTCGGGAAGCAGAGGCTCTCAAAAATATCCG 1682
90 aLcysGLy**LysLeuGlyArgPheProGlnValGlnAlaPhe..... 104
1681 TCGGCATATCCCTCAGTGAAGTGTGACACTGTACAGATGTTCGCGCGCA 1632
105 ...ValArgSerAspLysProLysLeuPheArgGly..... 115

```

```

1631 GTCATTAGATCATCATCGGAGACTGTCTCCGCGTCAGACTTATAT 1582
116 .....LeuGlnIleLysTyraValArgLysSerAspPro 126
1581 CATTTGGCTGAGCCATCATCATCAGTTGCTGACTGCTGATGC..... 1538
127 ValLeuLysLeuLeuAspAspAsnGlyAsnIleAlaGluGluLeuSerI 143
1537 .....TTGTCGCGT 1530
143 eleuLysTrpAsnThrAspSerValGluGluPheLeuSerGluLysLeu 159
1529 TGGCGCTTGGCGCAGCACTATCTACGAGATATCTCCACATCAAGTA 1481

seq_name: /cgn2_6/ptodata/2/lna/5A_COMB.seq:us-08-441-139-1

seq_documentation_block:
Sequence 1, Application US/08441139
Patent No. 5773245
GENERAL INFORMATION:
APPLICANT: Wiltrop, Dr. Karl D.
APPLICANT: Robinson, Anne S.
TITLE OF INVENTION: METHODS FOR INCREASING SECRETION OF
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESSES:
ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: NY
COUNTRY: USA
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/441,139
FILING DATE: 15-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/089,997
FILING DATE: 06-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: DIGILIO, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 8646
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-742-4343
TELEFAX: 516-742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2780 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 285..2333
US-08-441-139-1

alignment_scores:
Quality: 69.00      Length: 143
Ratio: 1.015        Gaps: 5
Percent Similarity: 47.552   Percent Identity: 23.776

alignment_block:
US-09-676-718a-1 x US-08-441-139-1 ..

```



```

TELEFAX: 212-705-5020
:
: TELEFAX: 212-705-5020
:
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
:   LENGTH: 35081 base pairs
:   TYPE: nucleic acid
:   STRANDEDNESS: single
:   TOPOLOGY: linear
:
: US-08-752-760A-1

alignment_scores:
  Quality: 67.00      Length: 94
  Ratio: 1.241        Gaps: 5
  Percent Similarity: 57.447  Percent Identity: 31.915

alignment_block:
US-09-676-718A-1 x US-08-752-760A-1/rev ..

Align seg 1/1 to reverse of: US-08-752-760A-1 from: 1 to: 35081

30 GYAlAGluphesererglualacysargluleugluphesereras 46
   ||| ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3750 GGCACAGATGCTCCCTCG..GGCAGATGGAAGTGGCAGACAGAGGGCT 3704
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
46 nleuLeu.CysSerSerCysAspLeuLeuGlylnpheasnleu..... 60
3703 ACTGCTGCTGCTGCGATGACCTCGAGCCAGCTGCTGCTGCTGCTATGG 3654
61 ..LeuGlnleuAspProAspCysArgGlyCysGlnleuGlnleuGlnleu 76
3653 CGGCTCGGGCTG...CCGGCGCTCGCGGGTGTCTCAACGACGAGATTC 3607
76 npheGluThrLysLysLeuTyraLagLysAlaLeuGlnleuValCysGly* 93
   : ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3606 CCCAGCGTGGCGGATGATGTCGAGGAAATTGCTGGAAGCAGCTGGGC. 3558
93 **LysLeuGlyArgPheProGlnValGlnAlaPheValArgSerAspLys 109
   ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3557 .....GCCCTCCACCGCTGCATCCCATCATCTGTGACGAACGCCGG 3516
110 ProlLysLeuPheArgGlyLeuGlnleuLys 119
3515 CCCATCTGGCAGCGCGGTTGTCAAAATTAAG 3486

seq_name: /cgn2_6/ptodata/2/1na/6A_COMB.seq:US-08-605-150A-3

seq_documentation_block:
: Sequence 3, Application US/08605150A
: Patent No. 6103520
: GENERAL INFORMATION:
: APPLICANT: Topfer, Reinhard
: APPLICANT: Hausmann, Ludwig
: APPLICANT: Schell, Jozsef
: TITLE OF INVENTION: GLYCEROL-3-PHOSPHATE DEHYDROGENASE
: NUMBER OF SEQUENCES: 17
: CORRESPONDENCE ADDRESS:
: ADDRESS: Klein & Szekeres
: STREET: 4199 Campus Drive, Suite 700
: CITY: Irvine
: STATE: CA
: COUNTRY: USA
: ZIP: 92715
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/605.150A
: FILING DATE: 01-MAR-1996
: CLASSIFICATION: 800

```

```

PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/EP94/02936
: FILING DATE: 02-SEP-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: DE P4329827.3
: FILING DATE: 03-SEP-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Szekeres, Gabor L.
: REGISTRATION NUMBER: 28, 675
: REFERENCE/DOCKET NUMBER: 542-04-PA
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 714-854-5502
: TELEFAX: 714-854-4897
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
:   LENGTH: 1464 base pairs
:   TYPE: nucleic acid
:   STRANDEDNESS: double
:   TOPOLOGY: linear
:   MOLECULE TYPE: cDNA to mRNA
:   HYPOTHETICAL: NO
:   ANTI-SENSE: NO
:   ORIGINAL SOURCE:
: ORGANISM: Cuphea lanceolata
: IMMEDIATE SOURCE:
: LIBRARY: ZAP cDNA library
: CLONE: C1GPDH109
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 45..1187
: US-08-605-150A-3

alignment_scores:
  Quality: 66.50      Length: 87
  Ratio: 1.255        Gaps: 5
  Percent Similarity: 60.920  Percent Identity: 33.333

alignment_block:
US-09-676-718A-1 x US-08-605-150A-3/rev ..

Align seg 1/1 to reverse of: US-08-605-150A-3 from: 1 to: 1464

20 AlaThrValLeuGlnAlaValSerAlaPheGlyAlaGluPheSerSerg1 36
   ||| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
974 GCCATCTTTGCAAAAGCCCTCAGCAGACTTCTCTGTTCTTCCGCCCAAC 925
36 uAlaCysArgGluLeuGlyPheSerSerAsnLeuLeuCysSerSergs. 52
   ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
924 AGTTGTGATGAGATCAGCAGCTCCGACGCTCGAAAAAAGTAGTGTC 875
53 .....AspLeuLeuGlyGln...PheAsnLeuLeuGlnleu 63
874 TTAACAGATGAAACAAACCTTGAGAGACGTTTCACTTCCCGCAGCC 825
64 AspPro...AspCysArgGlyCysGlnleuGlnleuGlnleuPheGluThr 79
   ||||| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
824 GATCTCATTAATGCGCGCTTTGTGTTGTTTCCCATCTCCAAATCCATCAA 775
79 rLys...LysLeuTyraLagLysAlaLeuLeuGlnleuValCysGly**LysL 95
   :||| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
774 CAAAACCCGCTGCTATGAGCCCAATATTTCTTCAAG...TGTCCACAAAGC 728
95 euGlyArgPhe 98
727 TCTACTCTCTTC 717

seq_name: /cgn2_6/ptodata/2/1na/6B_COMB.seq:US-08-942-686-1

seq_documentation_block:
: Sequence 1, Application US/08942686
: Patent No. 6183988
: GENERAL INFORMATION:

```

```

? APPLICANT: Bloch, Donald B.
? TITLE OF INVENTION: LEUKOCYTE-SPECIFIC PROTEIN AND GENE, AND
? TITLE OF INVENTION: METHODS OF USE THEREOF
? NUMBER OF SEQUENCES: 8
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: STERNER, KESSLER, GOLDSTEIN AND FOX, P.L.L.C.
? STREET: 1100 NEW YORK AVENUE, N.W. SUITE 600
? CITY: WASHINGTON
? STATE: D.C.
? COUNTRY: USA
? ZIP: 20005
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/942,686
? FILING DATE: 02-OCT-1997
? CLASSIFICATION: 514
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 60/027,347
? FILING DATE: 02-OCT-1996
? CLASSIFICATION: 514
? ATTORNEY/AGENT INFORMATION:
? NAME: GOLDSTEIN, JORGE A.
? REGISTRATION NUMBER: 29,021
? REFERENCE/DOCKET NUMBER: 0609.4330001/JAG/BJD
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 202-371-2600
? TELEFAX: 202-371-2540
? INFORMATION FOR SEQ ID NO: 1:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 2905 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: both
? TOPOLOGY: both
? MOLECULE TYPE: DNA (genomic)
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 107..2365
? US-08-942-686-1

alignment_scores:
  Quality: 66.50      Length: 133
  Ratio: 1.023        Gaps: 7
Percent Similarity: 48.872 Percent Identity: 24.060

alignment_block:
US-09-676-718A-1 x US-08-942-686-1 ..
Align seg 1/1 to: US-08-942-686-1 from: 1 to: 2905
34 SerSerGluAlaCysArgGluLeuGlyPheSerSerAsnLeuLeuCysSe 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1985 AGCCACAGTGTGTGACGAATCTGAGTCTGTGAGAGGACAGATGTGTC 2034
50 rSerCysAspLeuLeuGlyInPheAsnLeuLeuLeuAspProAspC 67
::: |||||:|||||:|||||:|||||:|||||:|||||:|||||:
2035 TGAGCAACAGTTGAATGTGACTTCTCTCTTGAAGTC..... 2074
67 ysArgGlyCysCysGlnGluAlaGlnPheGluThrLysLysLeuTyr 83
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2075 ....TATTGCGT...TCTGAGAGCTCCTTTTTCACAGATTCATC 2116
84 AlaGlyAlaIleLeuGluValCys..Gly**Lys.....Le 95
||| |||||:|||||:|||||:|||||:|||||:|||||:
2117 TATTATTATTATAGAGGCGGTCTCAAGCGCTGAAGGAGCCCATGTGT 2166
95 uGlyArg.....pHeProGlnValGlnA 103
|::: |||||:|||||:|||||:|||||:|||||:|||||:

```

```

2167 GGATAAATCAAGAAAGCTGATGACACGCTTACCCCCCAAGTGGAG 2216
103 lApheValArgSerAspLysProLysLeuPheArgGly..... 115
:::|||||:|||||:|||||:|||||:|||||:|||||:
2217 GGTGTGTACAA...GACATGGCGCTCATCTTCACAGAACACAGGCGCTCT 2263
116 .....LeuGlnIleLysTyr 120
|||||:|||||:|||||:|||||:|||||:|||||:
2264 TACAGATACAGAGATTTGGCCAAATGGGANTTTAGACTGGAGGCTGACTT 2313
120 rValArgGlySerAspProValIleLysLeuLeuAspAspAsnGlyAsn 136
: |||||:|||||:|||||:|||||:|||||:|||||:
2314 TGAGCAAGATTTCAAGAGAGTCTTGTCTATTCAGAAACAAATGGGAAC 2362
seq_name: /cgn2.6/prodata/2/ina/6B_comb.seq:US-09-180-439-1
seq_documentation_block:
? Sequence 1, Application US/09180439
? Patent No. 625532
? GENERAL INFORMATION:
? APPLICANT: Dixon, Mark S
? APPLICANT: Hatzixanthis, Kostas
? APPLICANT: Jones, David A
? APPLICANT: Jones, Jonathan DG
? TITLE OF INVENTION: Plant pathogen resistance genes and uses thereof
? FILE REFERENCE: 620 - 53
? CURRENT APPLICATION NUMBER: US/09/180,439
? CURRENT FILING DATE: 1998-12-06
? EARLIER APPLICATION NUMBER: PCT/GB97/01249
? EARLIER FILING DATE: 1997-05-08
? EARLIER APPLICATION NUMBER: GB 9609681.3
? EARLIER FILING DATE: 1996-05-09
? EARLIER APPLICATION NUMBER: GB 9619924.5
? EARLIER FILING DATE: 1996-09-24
? NUMBER OF SEQ ID NOS: 10
? SOFTWARE: Patentin Ver. 2.0
? SEQ ID NO 1
? LENGTH: 3979
? TYPE: DNA
? ORGANISM: Lycopersicon esculentum
? US-09-180-439-1

alignment_scores:
  Quality: 66.00      Length: 179
  Ratio: 0.857        Gaps: 7
Percent Similarity: 43.017 Percent Identity: 21.788

alignment_block:
US-09-676-718A-1 x US-09-180-439-1/rev ..
Align seg 1/1 to reverse of: US-09-180-439-1 from: 1 to: 3979
14 GlyLysArgLeuLeuAlaThrValLeu.....GlnAlaValSe 27
|||||:|||||:|||||:|||||:|||||:|||||:
2854 GGTTCCTCATTTGTTTATTCACACTGTCTCATTCCTTTCAAAATGTTCAA 2805
27 rAlaPheGlyAlaGluPheSerSerGluAlaCysArgGluLeuGlyPhe. 43
: :: |||||:|||||:|||||:|||||:|||||:|||||:
2804 TAGACTCGTGTGTAGTCTTTCGAGAAATGATTCGAGAGAGATCTATGA 2755
44 .....SerSerAsnLeuLeuCysSerSerCysAspLeuLeuGlyGln 57
|||||:|||||:|||||:|||||:|||||:|||||:
2754 TTCGAGATTCAGAAACATGATTCAGCCCTGATGATCTTATAGTCCA 2705
58 PheAsnLeuLeuGln..... 62
|||||:|||||:|||||:|||||:|||||:|||||:
2704 TGCATTTTATTCGATGTCAACCTTAAACTCTCAGCTCGCAAGTTC 2655
63 .....LeuAspProAspCysArgG 69
|||||:|||||:|||||:|||||:|||||:|||||:
2654 CAACCAATCGCAAAATGTGTGCTGTGAGTTGATTTGCTCTCTTAATCAAGA 2605

```


[illegible]

OM of: US-09-676-718A-1 to: N_Geneseq_032802.* out_format : pfs
Date: Sep 18, 2002 5:44 PM
About: Results were produced by the GenCore software, version 4.5.
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=framet_p2n.model -DEV=x1h
-O=/cgn2.1/USPTO/US09676718/rnat.17092002.143433.21344/app_query.fasta_1.222
-DB=N_Geneseq_032802 -OFMT=fastap -SUFFIX=ring -GAP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosome2
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=pcr
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NOM=EXT -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USFR=US09676718.ecgn1_135 -NCPU=6 -ICPU=3 -LONGLOG
-DEV=TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-676-718A-1
Query length: 162
Database: N_Geneseq_032802.*
Database sequences: 1736436
Database length: 858457221
Search time (sec): 201.290000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	! Documentation	..
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAA44484 +		819.00	1901.55	1.4e-97		
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AA234464 +		819.00	1893.70	3.9e-97		
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:AAK00680 +		819.00	1891.09	5.9e-97		
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAK16353 +		819.00	1890.56	5.9e-97		
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ43468 +		752.00	1734.15	3.0e-88		
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ43303 +		741.00	1719.77	1.9e-87		
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ43303 +		613.50	1418.66	1.1e-70		
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAA16434 +		552.50	1276.16	9.8e-63		
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABL13005 +		311.00	712.02	2.6e-3		
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:AAK63789 +		302.00	676.92	2.3e-2		
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABL13004 +		274.00	603.84	2.8e-2		
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAK34325 +		226.50	508.14	5.9e-20		
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAK34325 +		226.50	507.92	6.1e-20		
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAK53362 +		172.50	385.18	4.2e-1		
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAK03398 +		172.50	385.18	4.2e-1		
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAK30995 +		172.50	385.18	4.2e-1		
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAK13096 +		172.50	385.18	4.2e-1		
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAK78215 +		170.00	377.75	1.1e-12		
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAK78215 +		132.50	301.49	1.9e-0		
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAK18198 +		132.50	301.49	1.9e-0		
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAK44092 +		132.50	301.49	1.9e-0		
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAK50101 +		132.50	301.49	1.9e-0		
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAK14102 +		100.00	215.08	0.0012		
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAK22964 +		100.00	209.73	0.0025		
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:AAK06244 +		84.00	173.15	0.2689		
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAK99082 +		84.00	172.77	0.2822		
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAK85083 +		84.00	167.69	0.5817		
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABL14953 +		81.00	142.52	13.66		
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABL14952 +		81.00	132.60	48.77		
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAK76191 +		80.50	158.24	1.82		
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAK78001 +		80.50	158.24	1.82		
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAK28549 +		80.00	107.70	1.2e+0		
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:AAK64467 +		79.50	157.16	2.09		
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAK99480 +		79.50	155.15	2.70		
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAK22924 +		78.50	157.07	2.11		
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAK91633 +		78.00	58.81	5.7e+0		
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAK81459 +		77.00	158.05	1.86		
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABL27481 +		77.00	147.25	7.45		
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAK18646 +		77.00	136.80	28.44		
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABL27480 +		77.00	136.75	28.65		

/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:AAK91142 + 76.50 147.07 7.6
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAK00995 + 76.50 145.35 9.5
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAK02330 + 76.50 134.86 36.
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT:AAK55970 + 76.00 135.54 33.4
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAK91416 + 76.00 117.97 318

seq_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAA44484

seq_documentation_block:

ID AAA44484 standard; CDNA; 626 BP.

AC AAA44484;

DT 21-AUG-2000 (first entry)

DE Human secreted expressed sequence tag SEQ ID NO:1059.

KW Human; mouse; chicken; rat; secreted expressed sequence tag; SEST;
KW expressed sequence tag; EST; probe; chemotactic; proliferative;
KW immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;
KW thrombolytic; antiinflammatory; cyostatic; antibacterial; antifungal;
KW antiviral; antidiabetic; antiastrumatic; vulnary; antiparkinsonian;
KW antitumor; osteopathic; neuroprotective; nootropic; antipsoriatic;
KW cerebroprotective; anticonvulsant; antidepressant; gene therapy;
KW vaccine; autoimmune disorder; multiple sclerosis; allergic condition;
KW insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
KW lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
KW central nervous system disorder; Alzheimer's disease; stroke;
KW Parkinson's disease; Huntington's disease; coagulation disorder;
KW haemophilia; thrombosis; inflammatory disorder; Crohn's disease;
KW tumour; infection; depression; psoriasis; ss.

OS Homo sapiens.

PN WO200021991-A1.

XX 20-APR-2000.

PF 15-OCT-1999; 99WO-US24206.

PR 15-OCT-1998; 98US-0104436.

XX (GENY) GENETICS INST INC.

PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;

PT Metberg D, Treacy M, Bowman MR;

XX WPI: 2000-317938/27.

Isolated polynucleotides, and encoded proteins, comprising secreted expressed sequence tags (SESTs), useful for treating various disorders such as autoimmune, infectious, and central nervous system disorders - Claim 1; Page 474; 803pp; English.

AAA43426 to AAA45925 represent specifically claimed secreted expressed sequence tags (SESTs), isolated from human, mouse, chicken and rat tissue sources. The SESTs can have a range of activities depending on the tissues they were isolated from. The activities include:
chemotactic; proliferative; immunomodulatory; haematopoietic;
chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory;
cytostatic; antibacterial; antifungal; antiviral; antidiabetic;
antiastrumatic; vulnary; antitumor; osteopathic; neuroprotective;
nootropic; antiparkinsonian; antipsoriatic; cerebroprotective;
anticonvulsant; and antidepressant. The SESTs can be used for gene therapy and in vaccines. The SESTs are useful as probes for the identification and isolation of full-length cDNAs and genomic DNA molecules which correspond to the SESTs. Proteins encoded by the SESTs are useful in assays for determining biological activity and raising antibodies. In many cases they may be useful for treatment of autoimmune disorders (multiple sclerosis, insulin dependent diabetes), allergic conditions (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers, osteoporosis, osteoarthritis, central nervous system disorders

alignment_scores:

Quality: 819.00 Length: 162
 Ratio: 5.119 Gaps: 0
 Percent Similarity: 98.765 Percent Identity: 98.765

alignment_block:

US-09-676-718a-1 x AAX234464 ..

Align seg 1/1 to: AAX234464 from: 1 to: 1244

```

1 MetAlaIaGlyProSerGlyCysLeuValProAlaPheGlyLysArgIle 17
5 ATGGCGGCGTGGCCGAGTGGGTGTCGGTCCGCGCTTTGGCGTACGCTT 54
17 uLeuLeuAlaThrValLeuGlnAlaValSerAlaPheGlyAlaGluPheS 34
55 GTTGTGGGAGCTGTGCTTCAAGCGGTGTCTGCTTTGGGGAGAGATT 104
34 eSerGluAlaCysArgGluLeuGlyPheSerSerAsnLeuLeuCysSer 50
105 CATCGAGGAGCATGACAGAGATTAGGCTTTCTAGCAACTGCTTTCACG 154
51 SerCysAspLeuLeuGlyGlnPheAsnLeuGlnLeuAspProAspCy 67
155 TCTTGTGATCTTCTCGACAGTTCACACTGCTTCACTGATCTGATTTG 204
67 sArgGlyCysGlyGlnGluGluAlaGlnPheGluThrLysLysLeuTy 84
205 CAGAGGATGCTGTCAAGAGAACACAAATTTGAAACCAAAAGCTGTATG 254
84 IAGLYAlaIleLeuGluValGlyGly**LysLeuGlyArgPheProGln 100
255 CAGGAGCATTTCTTGAAGATTGTGATGAATAATGGGAAGGTTCCTCAA 304
101 ValGlnAlaPheValArgSerAspLysProLysLeuPheArgGlyLeu 117
305 GTCCAGACTTTCTTGAAGAGCATATAACCCAACTTTCAGAGGACTGCA 354
117 nIleLysTyValArgLysSerAspProValLeuLysLeuLeuAspAsp 134
355 AATCAAGATGTGCGTGGTTCAGACCTGTATTAAGCTTTGGAGCACA 404
134 sNGLYAsnIleAlaGluGluLeuSerIleLeuLysTrpAsnThrAspSer 150
405 ATGGGAACCTTCTGAAGAACTGAGCATTCTCAAAATGGAACACAGCACT 454
151 ValGluGluPheLeuSerGluLysLeuGluArgIle 162
455 CTGCAAGAAATTCCTGAGTCAAAAAGTTGGAAACCATTA 490

```

seq_name: /SIDS1/gcdata/hold.geneseq/geneseq-n.embl/NAL199. DAT: AAX00680

seq_documentation_block:

ID AAX00680 standard; DNA; 1563 BP.

AC AAX00680;

DT 25-MAR-1999 (first entry)

DE Human secreted protein gene 70 clone HMSGT42.

Human: secreted protein; fusion protein; gene therapy; protein therapy;
 diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
 developmental abnormality; fetal deficiency; blood; allergy; renal; ds;
 immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
 inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
 cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
 osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
 endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.

XX Homo sapiens.

XX OS

XX PN W09842738-A1.

XX 01-OCT-1998.

PD 19-MAR-1998; 98WO-US05311.

XX

PR 30-MAY-1997; 97US-0050937.

PR 21-MAR-1997; 97US-0041276.

PR 21-MAR-1997; 97US-0041277.

PR 21-MAR-1997; 97US-0041281.

PR 21-MAR-1997; 97US-0042344.

PR 30-MAY-1997; 97US-0048069.

PR 30-MAY-1997; 97US-0048094.

PR 30-MAY-1997; 97US-0048095.

PR 30-MAY-1997; 97US-0048096.

PR 30-MAY-1997; 97US-0048099.

PR 30-MAY-1997; 97US-0048131.

PR 30-MAY-1997; 97US-0048135.

PR 30-MAY-1997; 97US-0048154.

PR 30-MAY-1997; 97US-0048160.

PR 30-MAY-1997; 97US-0048186.

PR 30-MAY-1997; 97US-0048187.

PR 30-MAY-1997; 97US-0048188.

PR 30-MAY-1997; 97US-0048350.

PR 30-MAY-1997; 97US-0048351.

PR 30-MAY-1997; 97US-0048352.

PR 30-MAY-1997; 97US-0048353.

PR 05-AUG-1997; 97US-0054804.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

PI Brewer LA, Duan R, Edner R, Ferrie AM, Florence KA;

PI Greene JM, Hu JS, Latteur DM, Moore PA, Ni J, Olsen HS;

PI Rosen CA, Ruben SM, Shi Y, Young P;

DR MPI: 1999-070066/06.

XX P-PSDB: AAM67876.

XX

PT New isolated human genes and the secreted polypeptides they encode -

PT useful for diagnosis and treatment of e.g. cancers, neurological

PT disorders, immune diseases, inflammation or blood disorders

XX

PS Claim 1; Page 230-231; 385pp; English.

XX

This sequence represents a nucleic acid molecule which encodes a secreted human protein. The gene number, and the clone it is derived from, are detailed in the descriptor line. The gene can be used to generate fusion proteins by linking to the gene to a human immunoglobulin Fc portion (e.g. AAX00602) for increasing the stability of the fused protein as compared to the human protein only.

The invention relates to 87 novel genes and their fragments (nucleic acid sequences: AAX00611-X00724; amino acid sequences AAM67807-W68004) which are useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. Also, pathological conditions can be diagnosed by determining the amount of the new polypeptides in a sample or by determining the presence of mutations in the new polynucleotides. Specific uses are described for each of the 87 polynucleotides, based on which tissues they are most highly expressed in (see AAX00611 for described uses).

Sequence 1563 BP; 478 A; 252 C; 301 G; 531 T; 1 other.

alignment_scores:

Quality: 819.00 Length: 162
 Ratio: 5.119 Gaps: 0
 Percent Similarity: 98.765 Percent Identity: 98.765

alignment_block:

US-09-676-718a-1 x AAX00680 ..

Align seg 1/1 to: AAX00680 from: 1 to: 1563

1 MetAlaIaGlyProSerGlyCysLeuValProAlaPheGlyLysArgIle 17

```

|||||
40 ATGGCGGCTGGCCGAGTGGTGTGTCGCCGCGTTGGGCTACAGCTT 89
17 uLeuLeuAlaThrValLeuGlnAlaValSerAlaPheGlyAlaGluPheS 34
|||||
90 GTTGTGGCGACTGTGCTTCAAGCGTGCTGTGCTTTGGGCGAGGTTT 139
34 eSerGluAlaCysArgGluLeuGluPheSerSerAsnLeuLeuCysSer 50
|||||
140 CATCGAGGAGCATGCGAGAGTGAAGCTTTTCTTACCACTTCTTGGCAGC 189
51 SerCysAspLeuLeuGlyGlnPheAsnLeuLeuGlnLeuAspProAspCy 67
|||||
190 TCTTGTGATCTTCTCGGACATGTTCACTGCTTCACTGATGCTGATG 239
67 sArgGlyCysCysGlnGluGlnAlaGlnPheGluThrLysLysLeuTYRA 84
|||||
240 CAGAGGATGCTGTGAGGAGCAACAATTTGAAACCAAAAGCTGTATG 289
84 laGlyAlaIleLeuGluValCysGly**LysLeuGlyArgPheProGln 100
|||||
290 CAGAGGATGCTGTGAGGAGCAACAATTTGAAACCAAAAGCTGTATG 339
101 ValGlnAlaPheValArgSerAspLysProLysLeuPheArgGlyLeuG 117
|||||
340 GTCCACGCTTTTGTAGAGATGAACCCAACTGTTCAAGAGCACTGCA 389
117 nIleLysTyValAlaArgLysSerAspProValLeuLysLeuLeuAspA 134
|||||
390 AATCAAGATATGCTCGGCTGTCAGACCCGTATTAAGCTTTTGGACGCA 439
134 snGlyAsnIleAlaGlnGluLeuSerIleLeuLysTrpAsnThrAspSer 150
|||||
440 ATGGGAACATGCTGAAGAACTGAGCAATTCCAATGGAACACAGACAGT 489
151 ValGluGluPheLeuSerGluLysLeuGluArgIle 162
|||||
490 GTAGAGAAATTCCTGAGTGAAGAACTTGAACGCA 525
seq_name: /STDS1/gcgdata/mol-d-geneseq/geneseqn-emb1/NA2000.DAT:AAF16335
seq_documentation_block:
ID AAF16335 standard; cDNA; 1637 BP.
AC AAF16335;
XX
XX
DT 13-MAR-2001 (first entry)
XX
XX
DE Human prostate cancer antigen nucleotide sequence SMO ID NO:770.
XX
XX
KW Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
KW neuroprotective; cytosolic; cardiovascular; immunomodulatory; muscular;
KW vulnerrary; gastrointestinal; nephrotoxic; antiinfective; gynaecological;
KW antibacterial; gene therapy; neural; immune; reproductive; renal;
KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
XX
XX
wound; infectious disease; ss.
OS Homo sapiens.
XX
XX
PN WO20005174-A1.
XX
XX
PD 21-SEP-2000.
XX
XX
PF 08-MAR-2000; 2000MO-US05988.
XX
XX
PR 12-MAR-1999; 99US-0124270.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX
PA (ROSE/) ROSEN C A.
XX
XX
PI Rosen CA, Ruben SM;
XX
XX
DR WPI; 2000-587513/55.

```

```

DR P-PSDB; AAB57132.
XX
XX Prostate cancer associated gene sequences, referred to as prostate
PT cancer antigens, useful for treatment, prevention, and diagnosis of
PT disorders such as prostate cancer
XX
XX
PS Claim 1; Page 1189-1190; 2338pp; English.
XX
XX AAF15566 to AAF16505 encode the human prostate cancer associated
CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
CC The prostate cancer antigens can have neuroprotective, cytosolic,
CC cardioactive, immunomodulatory, muscular, vulnerrary, gastrointestinal,
CC nephrotoxic, antiinfective, gynaecological and antibacterial activities,
CC and can be used in gene therapy. The prostate cancer antigen
CC polynucleotides may be used for detection of prostate cancer, chromosome
CC identification, as chromosome markers, and for numerous other diagnostic
CC or research purposes. The prostate cancer antigens may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to
CC AAB57303 represent sequences used in the exemplification of the present
CC invention.
XX
XX
SO Sequence 1637 BP; 500 A; 273 C; 323 G; 541 T; 0 other;

```

```

alignment_scores:
  Quality: 819.00      Length: 162
  Ratio: 5.119        Gaps: 0
Percent Similarity: 98.765 Percent Identity: 98.765

```

alignment_block:

US-09-676-718A-1 x AAF16335 ..

Align seg 1/1 to: AAF16335 from: 1 to: 1637

```

1 MetAlaAlaGlyProSerGlyCysLeuValProAlaPheGlyLysArgIle 17
|||||
99 ATGGCGGCTGGCCGAGTGGTGTGTCGCCGCGCTTGGGCTACAGCTT 148
17 uLeuLeuAlaThrValLeuGlnAlaValSerAlaPheGlyAlaGluPheS 34
|||||
149 GTTGTGGCGACTGTGCTTCAAGCGTGCTGTGCTTTGGGCGAGGTTT 198
34 eSerGluAlaCysArgGluLeuGluPheSerSerAsnLeuLeuCysSer 50
|||||
199 CATCGAGGAGCATGCGAGAGTGAAGCTTTCTTACCACTTCTTGGCAGC 248
51 SerCysAspLeuLeuGlyGlnPheAsnLeuLeuGlnLeuAspProAspCy 67
|||||
249 TCTTGTGATCTTCTCGGACATGTTCACTGCTTCACTGATGCTGATG 298
67 sArgGlyCysCysGlnGluGlnAlaGlnPheGluThrLysLysLeuTYRA 84
|||||
299 CAGAGGATGCTGTGAGGAGCAACAATTTGAAACCAAAAGCTGTATG 348
84 laGlyAlaIleLeuGluValCysGly**LysLeuGlyArgPheProGln 100
|||||
349 CAGAGGATGCTGTGAGGAGCAACAATTTGAAACCAAAAGCTGTATG 398
101 ValGlnAlaPheValArgSerAspLysProLysLeuPheArgGlyLeuG 117
|||||
399 GTCCACGCTTTTGTAGAGATGAACCCAACTGTTCAAGAGCACTGCA 448
117 nIleLysTyValAlaArgLysSerAspProValLeuLysLeuLeuAspA 134
|||||
449 AATCAAGATATGCTCGGCTGTCAGACCCGTATTAAGCTTTTGGACGCA 498
134 snGlyAsnIleAlaGlnGluLeuSerIleLeuLysTrpAsnThrAspSer 150
|||||
499 ATGGGAACATGCTGAAGAACTGAGCAATTCCAATGGAACACAGACAGT 548
151 ValGluGluPheLeuSerGluLysLeuGluArgIle 162

```


XX (GSET) GENSET.
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX WPI; 2000-038446/03.
DR P-PSDB; AAY65395.
XX
XX Novel secreted protein 5' expressed sequence tag sequences used in
PT diagnostic, forensic, gene therapy, and chromosome mapping procedures
PS
PS Claim 1; Page 570; 837pp; English.
XX
XX AA242265 to AA243075 represent novel 5' expressed sequence tag (EST)
CC sequences, corresponding to human secreted proteins. AAY64651 to
CC AAY6538 represent the EST-related proteins corresponding to AA242265 to
CC AA243052. The 5' ESTs can be used for producing secreted human gene
CC products. They can be used to identify and isolate 5' untranslated
CC regions (UTRs) and upstream regulatory regions which control the
CC location, development stage, rate, and quantity of protein synthesis,
CC well as stability of mRNA. The ESTs are also useful as probes for
CC chromosome mapping, and to obtain full length cDNA clones. The ESTs can
CC also be used in forensic procedures to identify individuals, or in
CC diagnostic procedures to identify individuals having genetic diseases
CC resulting from abnormal gene expression. The products may also be used i
CC gene therapy protocols. The nucleic acids encoding signal peptides can b
CC used for directing extracellular secretion of a polypeptide or the
CC insertion of a polypeptide into a membrane, or importing a polypeptide
CC into a cell. The proteins encoded by the EST sequences may be useful in
CC treating a variety of human conditions. Secreted proteins have
CC therapeutic value, and the identification of new secreted proteins is
CC valuable. AA242249 to AA24264 and AAY64644 to AAY64650 represent
CC sequences used in the exemplification of the present invention.
SQ
SQ Sequence 673 BP; 161 A; 130 C; 193 G; 188 T; 1 other;

alignment_scores:
Quality: 742.00 Length: 146
Ratio: 5.153 Gaps: 0
Percent Similarity: 98.630 Percent Identity: 98.630

alignment_block:
US-09-676-718A-1 x AA243009 ..

Align seg 1/1 to: AA243009 from: 1 to: 673

1 MetaIaIaIaGlyProSerGlyCysLeuValProIaIaPheGlyLysArgLe 17
|||||
227 ATGGCGGCGTGGCGCAGATGGGTGTCTGTGTCGGCGCTTTGGGCTACCGGTT 276
|||||
17 uLeuLeuIaArthrValLeuGlnIaIaValSerAlaPheGlyIaIaIuPheS 34
277 GTTGTGGCGCACTGTGCTTCAAGCGGTGTCTGCTTTTGGGCGACAGATTTT 326
|||||
34 eSerGlnIaIaCysArgGluLeuGlyPheSerSerAsnLeuLeuCysSer 50
|||||
327 CATCGGAGGCAATGCAGAGAGATTAGCGTTTCTTACCACTGCTTTGGAGC 376
|||||
51 SerCysAspLeuLeuGlyGlnPheAsnLeuLeuGlnLeuAspProAspCy 67
377 TCTTGTGATCTTCTCGGACATTCACCTGCTTACGCTGGATCTCTGATTG 426
|||||
67 sArgGlyCysCysGlnGluGlnIaIaGlnPheGluThrLysLysLeuTyra 84
|||||
427 CAGAGGATGTCTGTAGGAGGAAGCACAATTGAAACCAAAAAGCTGATG 476
|||||
84 IaGlyAlaIleLeuGluValCysGly**LysLeuGlyAArgPheProGln 100
|||||
477 CAGGAGCATATCTTCAAGCTTGTGGATCAAAATTTGGGAGAGTCTCCCAA 526
|||||
101 ValGlnIaIaPheValArgSerAspLysProLysLeuPheArgGlyLeuGln 117
527 GTCCAGGCTTTGTGTAGGATGATTAACCCAAACTGTTCAAGAGGACTGCA 576

```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ43003
seq_documentation_block:
ID      AAZ43003 standard; cDNA; 623 BP.
XX
AC      AAZ43003;
XX
DE      01-FEB-2000 (first entry)
XX
Human 5' EST isolated from a cDNA library SEQ ID NO:762.
XX
Human: 5' EST: expressed sequence tag; secreted protein; diagnosis;
KW      gene therapy; chromosome mapping; upstream regulatory sequence;
KW      forensic; location; development; protein synthesis; stability;
KW      regulation; identification; ss.
XX
OS      Homo sapiens.
XX
PN      MO9953051-A2.
XX
PD      21-OCT-1999.
XX
PF      09-APR-1999; 99WO-IB00712.
XX
PR      09-APR-1998; 98US-0057719.
XX      28-APR-1998; 98US-0069047.
XX
PA      (GEST ) GENSET.
XX
PI      Dumas Mline Edwards J, Duclert A, Giordano J;
PI      WPI; 2000-038446/03.
DR      P-PSDB; AAY65389.
XX
PT      Novel secreted protein 5' expressed sequence tag sequences used in
PT      diagnostic, forensic, gene therapy, and chromosome mapping procedures
XX
PS      Claim 1; Page 566-567; 837pp; English.
XX
AAZ42265 to AAZ43075 represent novel 5' expressed sequence tag (EST)
sequences, corresponding to human secreted proteins. AAY64651 to
AAV65438 represent the EST-related proteins corresponding to AAZ42265 to
AAZ43052. The 5' ESTs can be used for producing secreted human gene
products. They can be used to identify and isolate 5' untranslated
regions (UTRs) and upstream regulatory regions which control the
location, development stage, rate, and quantity of protein syntheas, as
well as stability of mRNA. The ESTs are also useful as probes for
chromosome mapping, and to obtain full length cDNA clones. The ESTs can
also be used in forensic procedures to identify individuals, or in
diagnostic procedures to identify individuals having genetic diseases
resulting from abnormal gene expression. The products may also be used
in gene therapy protocols. The nucleic acids encoding signal peptides can
be used for directing extracellular secretion of a polypeptide or the
insertion of a polypeptide into a membrane, or importing a polypeptide
into a cell. The proteins encoded by the EST sequences may be useful in
treating a variety of human conditions. Secreted proteins have
therapeutic value, and the identification of new secreted proteins is
valuable. AAZ42249 to AAZ42264 and AAY64644 to AAY64650 represent
sequences used in the exemplification of the present invention.
XX
Sequence 623 BP; 143 A; 120 C; 182 G; 175 T; 3 other;
alignment_scores:

```

Quality:	613.50	Length:	147
Ratio:	4.869	Gaps:	2
Percent Similarity:	85.714	Percent Identity:	85.034

alignment_block:

US-09-676-718A-1 x AAZ43003

Align seg 1/1 to: AAZ43003 from: 1 to: 623

1	MetAlaIaIaGlyProSerGlyCysLeuValProAlaPheGlyLysArg	17
2		
3		
4		
5		
6	227 ATGGCGGCTGGGCGGAGTGGGTGTGTGGTCCGGCGCTTTGGGCTACGGTT	276
7	17 ULeuLeuAlaThrValLeuGlnAlaValSerAlaPheGlyAlaGluPheS	34
8		
9	277 GTTGTGGCGCAGCTGGCTTCAAGCGGTGTCTCTTTGGCGCAGAGTTT	326
10	34 eSerGlyAlaCysArgGluLeuGlyPheSerSerAsnLeuLeuGlySer	50
11		
12	327 CATCGGAGCATTGCAAGAGATTGGCTTTTCTAGCAACTGGCTTTCAC	376
13	51 SerCysAspLeuLeuGlyGlnPheAsnLeuLeuGlnLeuAspProAsp	67
14		
15	377 TCTGTGATCTTCTCGGACAGTCAACCTGGCTCAGCTGATCCTGATTG	426
16	67 sArgGlyCysGlyGlnGluGlnAlaGlnPheGluThrLysLysLeuTyr	84
17	427 CAGGGAGATGGTGTCAAGGAGAAACCAATTGTAAACCAAAAAGCTGTATG	476
18	84 IaGlyAlaIaIaLeuGlnValCysGly**LysLeuGlyArgPheProGln	100
19		
20	477 CAGGAGCATATCTTGAATTTGTGTGATGAATAATTGGGAAGGTTCCCT	523
21	101 ValGlnAlaPheValArgSerAspLysProLysLeuPheArgGlyLeuG	117
22		
23	524CA	525
24	117 nIleLysTyrValArgLysSerAspProValLeuLysLeuLeuAsp	133
25		
26	526 AGTCAAGATGATGTCGGTGGTTCAGACCCGTATTAAAGCTTTTGGACGAC	575
27	134 AsnGlyAsnIleIaGluGlnLeuSerIleLeuLysTrp	146
28		
29	576 AATTGGAAACATTTGGTGAAGAACTGGACATTCTTCCAATGG	614

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAA16434

seq_documentation_block:

ID	AAA16434	standard; DNA; 578 BP.
...		

AC	AAA16434;
XX	
DT	14-JUN-2000 (first entry)
XX	
DE	Human colon cancer differentially expressed nucleotide sequence #439
XX	
KX	Colon cancer; detect; differential expression; human; treatment;
KW	detect mutation; non-invasive diagnostic method; ds.
XX	
OS	Homo sapiens.
XX	
PN	MO200012702-A2.
PN	
PD	09-MAR-2000.
XX	
PF	30-AUG-1999; 99MO-US19424.
XX	
PR	31-AUG-1998; 98US-0098639.
PR	27-JAN-1999; 99US-0117393.
XX	
PA	(FARB) BAYER CORP.
XX	
TI	Endege WO, Steinmann KE, Astle JH, Burgess CC, Carroll E;

PI Catino TJ, Dwaivedi P, Ford DM, Lewis ME, Molino GA, Monahan JE,
PI Schlegel R;
XX
DR WPI; 2000-256641/22.

PT Novel nucleic acids and proteins for identifying therapeutic agents
PT useful for treating and diagnosing cancer, especially colon cancer
XX
PS Claim 16; Page 301-302; 345pp; English.

CC This sequence represents a human nucleotide sequence which is
CC differentially expressed in colon cancer cells compared to the expression
CC levels in normal cells. The nucleotide sequence can be used as a source
CC of primers and probes. The nucleotide sequence is useful for determining
CC the phenotype of a cell by detecting the differential expression of the
CC sequence relative to a normal cell. The probes derived from the sequence
CC can also be used to determine the phenotype of cells in a sample. Probes
CC and antibodies which hybridise to the nucleotide sequence can also be
CC used to determine the phenotype of a cell. The primers are useful for
CC detecting a mutation in a test nucleotide sequence and also for detecting
CC cancer, preferably colon cancer. Antibodies against the protein encoded
CC by the nucleotide sequence can also be used in a method to detect colon
CC cancer. The diagnostic method is non-invasive and accurate for diagnosing
CC colon cancer at an early stage.

SQ Sequence 578 BP; 143 A; 102 C; 152 G; 160 T; 21 other;

alignment_scores:

Quality:	552.50	Length:	136
Ratio:	4.604	Gaps:	3
Percent Similarity:	88.235	Percent Identity:	86.029

```
alignment_block:
```

US-09-676-718A-1 X AAA16434

Align seg 1/1 to: AAA16434 from: 1 to: 578

```

1 MetAlaIaIaGlyProSerGlyCysLeuValProAlaPheGlyLysArgIle 17
43 ATGGCGGCTGGGCCGAGTGGGTCTTGCTGCCGGCGCTTTGGGCTACGGTT 92
17 uLeuLeuAlaThrValLeuGlnAlaValSerAlaPheGlyAlaGluPheS 34
93 GTTGTGGCGACTGTGCTTCAAGCGGTGCTCTTTGGGCGAGAGTTT 142
34 eArgGlyAlaCysArgGluLeuGlyPheSerSerAsnLeuLeuCysSer 50
143 CATCGGAGGCGATGCAAGAGATTAGGCTTTTCAGCAACTGTGCTTACAGC 192
51 SerCysAspLeuLeuGlyGlnPheAsnLeuLeuGlnLeuAspProAspCy 67
193 TCTTGTGATCTTCTCGGACAGTTCAACTGCTTACGCTGATCTGATTG 242
67 sArgGlyCysCysGlnGluGlnAlaGlnPheGluThrLysLysLeuTYR 84
243 CAAAGGATGCTGTACGAGGAAAGCAACTTTTCAACCAAAAAGCTGTATG 292
84 IaGlyAlaIleLeuGlnValCysGly**LysLeuGlyrArgPheProGln 100
293 CAGGAGCTATTCTTGAAGTTGNGSAGTGAATAATTGGGAAGGTTCCTTMA 342
101 ValGlnAlaPheValIaArgSerAspLysProLysLeuPhe...ArgGlyYL 116
343 GTCCAAACTTTTGTATACAGTGAATAAACCCCAAACTGTTTCAGAAAGACTG 392
116 euGlnIleLysTYrValArgGlySerAspProValLeuLeuLeuLeuAs 132
393 CCAATTAAGTATCTNNCGTGTTCACCAACCTGAATTAAAAAGCTTTT 442
132 pAsp 133
443 NGAC 446

```

```
seq_name: /STDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT:ABL13005
seq_documentation_block:
ID   ABL13005 standard; cDNA; 429 BP.
XX
XX   ABL13005;
XX
XX   26-MAR-2002 (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 33497.
XX
XX   Drosophila; developmental biology; cell signalling; insecticide;
XX   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
XX   WO200171042-A2.
XX
XX   27-SEP-2001.
XX
XX   23-MAR-2001; 2001WO-0509231.
XX
XX   23-MAR-2000; 2000US-191637P.
XX
XX   11-JUL-2000; 2000US-0614150.
XX
XX   (PEKE ) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
XX
XX   WPI: 2001-656860/75.
XX
XX   P-PSDB; ABB68902.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
XX   Claim 1; SEQ ID NO 33497; 21bp + Sequence Listing; English.
XX
XX   The invention relates to an isolated nucleic acid detection reagent
XX   capable of detecting 1000 or more genes from Drosophila. The invention is
XX   useful in developmental biology and in elucidating cell signalling and
XX   cell-cell interactions in higher eukaryotes for the development of
XX   insecticides, therapeutics and pharmaceutical drugs. The invention
XX   discloses genomic DNA sequences (ABL1840-ABL16175) and the encoded proteins
XX   (ABB5737-ABB72072).
XX   (ABB5737-ABB72072).
XX   The sequence data for this patent did not form part of the printed
XX   specification, but was obtained in electronic format directly from WIPO
XX   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 429 BP; 118 A; 103 C; 119 G; 89 T; 0 other;
XX

alignment_scores:
Quality: 311.00      Length: 114
Ratio: 3.659        Gaps: 0
Percent Similarity: 74.561      Percent Identity: 51.754

alignment_block:
US-09-676-718A-1 x ABL13005 ..

Align seg 1/1 to: ABL13005 from: 1 to: 429

48 LeucySerSerCysAspLeuLeuGlyGlnPheAsnLeuLeuGlnLeuAs 64
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1 ATGTTCTCCAGTTCGAAAACTGGATTCGATTCGATTCGATTCACATCAA 50
64 PTrAspCysArgGlyCysGlyGlnGluAlaGlnPheGluThrTyrSL 81
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::
51 GCCTCAGTGTAAAGCAATGCTGCACTTGTGATCAGCAGCGCGCAGCAGC 100
81 ySLeuTyrAlaGlyAlaIleLeuGluValCysGly**LysLeuGlyArg 97
```

```
seq_name: /STDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAF63789
seq_documentation_block:
ID   AAF63789 standard; cDNA; 1453 BP.
XX
XX   AAF63789;
XX
XX   03-APR-2001 (first entry)
XX
XX   Human secreted protein gene 1 SEQ ID NO:11.
XX
XX   Human; immunosuppressive; antiarthritic; antirheumatic; noctropic;
XX   antiproliferative; cytosolic; cardiant; vasotropic; cerebroprotective;
XX   neuroprotective; antibacterial; virucide; fungicide; ophtalmological;
XX   vulnerary; autoimmune disease; hyperproliferative disorder; cancer;
XX   cardiovascular disorder; cerebrovascular disorder; infection;
XX   nervous system disorder; ocular disorder; chemotaxis; food additive;
XX   secreted protein; ss.
XX
XX   Homo sapiens.
XX
XX   WO200077021-A1.
XX
XX   21-DEC-2000.
XX
XX   01-JUN-2000; 2000WO-US15135.
XX
XX   11-JUN-1999; 99US-0138632.
XX
XX   (HUMA-) HUMAN GENOME SCI INC.
XX   (ROSE/) ROSEN C A.
XX
XX   Rosen CA, Ruben SM, Komatsoulis GA;
XX
XX   WPI: 2001-071257/08.
XX
XX   P-PSDB; AAB75240.
XX
XX   Nucleic acid molecules encoding human secreted proteins, used in
XX   preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
XX   Parkinson's diseases and cancers -
XX
XX   Claim 1; Page 440; 530pp; English.
XX
XX   This invention relates to polynucleotide sequences AAF63789 - AAF63836
XX   which encode human secreted proteins AAB75260 - AAB75287. Included in the
XX   invention are protein sequences AAB75288 - AAB75341 which are fragments
XX   of the secreted proteins and amino acid sequences with which these
XX   fragments share homology. Examples of the activities of the proteins and
XX   polynucleotides and the activities of their agonists and antagonists
XX   include; immunosuppressive; antiarthritic; antirheumatic;
XX   antiproliferative; cytosolic; cardiant; vasotropic; cerebroprotective;
XX   noctropic; neuroprotective; antibacterial; virucide; fungicide;
XX   ophtalmological; and vulnerary activity. The protein and polynucleotide
XX   sequences, their agonists and antagonists may be useful for treating,
```

CC preventing and diagnosing diseases and disorders such as autoimmune
CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders
CC e.g. neoplasms of the breast or liver, cardiovascular disorders
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC angogenesis, nervous system disorders e.g. Alzheimer's disease,
CC infections caused by bacteria, viruses and fungi and ocular disorders
CC e.g. corneal infection. The polypeptides can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities. Included in the invention are
CC oligonucleotides AAF63780-AAF63788 and peptide AAB75239 which are used
CC in the identification and characterisation of the DNA and protein
CC sequences of the invention.

SO Sequence 1453 BP; 473 A; 209 C; 229 G; 542 T; 0 other;

alignment_scores:
Quality: 302.00 Length: 60
Ratio: 5.033 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-676-718a-1 x AAF63789 ..

Align seg 1/1 to: AAF63789 from: 1 to: 1453

103 AlaphValaIarSerAspProValleuLysLeuAspAspans1ya 119
|||
226 GCTTTGTTGTAAGAGTGTATTAACCAACTGTTCAGAGACTGCAATCA 275
|||
119 sTyrValaIarGlySerAspProValleuLysLeuAspAspans1ya 136
|||
276 GTATGTCCTGCGTTCAGACCTGTATTAAGCTTTTGACAGACATGGA 325
|||
136 snlleAlaIugluLeuSerIleuLysTrpAsnThrAspSerValGlu 152
|||
326 ACATTCCTGTAAGAACTGACGATTCCTCAATAGAAACAGACAGCTGTGAA 375
|||
153 GluPheLeuSerGluLysLeuGluArgIle 162
|||
376 GAATTCCTGAGTGAATAAGTTGGAACGCATTA 405
|||

seq_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABLI3004

seq_documentation_block:

ID ABLI3004 standard; CDNA; 2746 BP.

XX ABLI3004;

DE 26-MAR-2002 (first entry)

XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 33494.

KM Drosophila; developmental biology; cell signalling; insecticide;

XX pharmaceutical; gene; ss.

OS Drosophila melanogaster.

XX WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

PA (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PMD, Myers EM;

XX WPI; 2001-656860/75.
DR P-PSDB; ABB68901.

PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -

XX Claim 1; SEQ ID NO 33494; 21bp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABLI6176-ABLI30511), expressed DNA
CC sequences (ABLI01840-ABLI6175) and the encoded proteins
CC (ABBI57737-ABBI2072).

CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

SO Sequence 2746 BP; 853 A; 516 C; 546 G; 831 T; 0 other;

alignment_scores:
Quality: 274.00 Length: 97
Ratio: 3.753 Gaps: 0
Percent Similarity: 75.258 Percent Identity: 53.608

alignment_block:
US-09-676-718a-1 x ABLI3004/reverse ..

Align seg 1/1 to reverse of: ABLI3004 from: 1 to: 2746

65 ProAspPysArgGlyCysGlnGluAlaIlePheGluTrpLysIle 81
|||
1378 CCTCAGGTGAAGCAATGCTGCACTTGTGATCAGACGCGCGGACAGCG 1329
|||
81 sLeuTyrAlaGlyAlaIleLeuGluValCysGly**IysLeuGlyArgp 98
|||
1328 GACATATGCCAAGCAATTCCTGAGGTGTGTGACCTGCMAATTCGGGCGCT 1279
|||
98 heProGluValGlnAlaPheValaIarSerAspLysProLysLeuPheArg 114
|||
1278 ATCCGACAGATTGACGCTTATTCAAAGCGCGGACCTGCCAAGTTCGCC 1229
|||
115 GlyLeuGlnIleLysTyrValaIarGlySerAspProValleuLysLeu 131
|||
1228 AACCTGCAGATCAAAATACGTAAAGGAGCTGATCCGTTGTTAAGCTCCT 1179
|||
131 uAspAspAnsGlyAsnIleAlaIugluLeuSerIleuLysTrpAsnT 148
|||
1178 AGATGCCAGTGGCAAAAGTGCAGAGACGTTGTCATTAACCAAGTGAACA 1129
|||
148 hrAspSerValGluGluPheLeuSerGluLysLeuGluArg 161
|||
1128 CAGACACTGTGAGAGAGTCTTCGAAACGCAATCTGGCCAAAG 1088
|||

seq_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AA47500

seq_documentation_block:

ID AA47500 standard; DNA; 682 BP.

XX AA47500;

DE 18-OCT-2000 (first entry)

XX Arabidopsis thaliana DNA fragment SEQ ID NO: 54045.

KM Hybridisation assay; genetic mapping; gene expression control;

KM protein identification; signal transduction pathway;

KM metabolic pathway; promoter; termination sequence; ss.

XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
PD
XX 25-FEB-2000; 2000EP-0301439.
PF
XX
XX 25-FEB-1999; 9905-0121825.
PR 05-MAR-1999; 9905-01233180.
PR 09-MAR-1999; 9905-0123548.
PR 23-MAR-1999; 9905-0125788.
PR 25-MAR-1999; 9905-0126284.
PR 29-MAR-1999; 9905-0126785.
PR 01-APR-1999; 9905-0127462.
PR 06-APR-1999; 9905-0128242.
PR 08-APR-1999; 9905-0128714.
PR 16-APR-1999; 9905-0129845.
PR 19-APR-1999; 9905-0130077.
PR 21-APR-1999; 9905-0130449.
PR 23-APR-1999; 9905-0130510.
PR 28-APR-1999; 9905-0130891.
PR 30-APR-1999; 9905-0132048.
PR 30-APR-1999; 9905-0132407.
PR 04-MAY-1999; 9905-0132484.
PR 05-MAY-1999; 9905-0132485.
PR 06-MAY-1999; 9905-0132486.
PR 06-MAY-1999; 9905-0132487.
PR 07-MAY-1999; 9905-0132863.
PR 11-MAY-1999; 9905-0134256.
PR 14-MAY-1999; 9905-0134218.
PR 14-MAY-1999; 9905-0134219.
PR 14-MAY-1999; 9905-0134221.
PR 18-MAY-1999; 9905-0134370.
PR 19-MAY-1999; 9905-0134768.
PR 20-MAY-1999; 9905-0135124.
PR 21-MAY-1999; 9905-0135353.
PR 24-MAY-1999; 9905-0135629.
PR 25-MAY-1999; 9905-0136021.
PR 27-MAY-1999; 9905-0136392.
PR 28-MAY-1999; 9905-0136782.
PR 01-JUN-1999; 9905-0137222.
PR 03-JUN-1999; 9905-0137528.
PR 04-JUN-1999; 9905-0137502.
PR 07-JUN-1999; 9905-0137724.
PR 08-JUN-1999; 9905-0138094.
PR 10-JUN-1999; 9905-0138540.
PR 10-JUN-1999; 9905-0138847.
PR 14-JUN-1999; 9905-0139119.
PR 16-JUN-1999; 9905-0139452.
PR 16-JUN-1999; 9905-0139453.
PR 17-JUN-1999; 9905-0139492.
PR 18-JUN-1999; 9905-0139454.
PR 18-JUN-1999; 9905-0139455.
PR 18-JUN-1999; 9905-0139456.
PR 18-JUN-1999; 9905-0139457.
PR 18-JUN-1999; 9905-0139458.
PR 18-JUN-1999; 9905-0139459.
PR 18-JUN-1999; 9905-0139460.
PR 18-JUN-1999; 9905-0139461.
PR 18-JUN-1999; 9905-0139462.
PR 18-JUN-1999; 9905-0139463.
PR 18-JUN-1999; 9905-0139750.
PR 18-JUN-1999; 9905-0139763.
PR 21-JUN-1999; 9905-0139817.
PR 22-JUN-1999; 9905-0139839.
PR 23-JUN-1999; 9905-0140353.
PR 23-JUN-1999; 9905-0140354.
PR 24-JUN-1999; 9905-0140695.
PR 28-JUN-1999; 9905-0140823.

PR 29-JUN-1999; 9905-0140991.
PR 30-JUN-1999; 9905-0141287.
PR 01-JUL-1999; 9905-0141842.
PR 01-JUL-1999; 9905-0142154.
PR 02-JUL-1999; 9905-0142055.
PR 06-JUL-1999; 9905-0142390.
PR 08-JUL-1999; 9905-0142803.
PR 09-JUL-1999; 9905-0142920.
PR 12-JUL-1999; 9905-0142977.
PR 13-JUL-1999; 9905-0143542.
PR 14-JUL-1999; 9905-0143624.
PR 15-JUL-1999; 9905-0144005.
PR 16-JUL-1999; 9905-0144085.
PR 16-JUL-1999; 9905-0144086.
PR 19-JUL-1999; 9905-0144325.
PR 19-JUL-1999; 9905-0144331.
PR 19-JUL-1999; 9905-0144332.
PR 19-JUL-1999; 9905-0144333.
PR 19-JUL-1999; 9905-0144334.
PR 19-JUL-1999; 9905-0144335.
PR 20-JUL-1999; 9905-0144352.
PR 20-JUL-1999; 9905-0144632.
PR 21-JUL-1999; 9905-0144884.
PR 21-JUL-1999; 9905-0144814.
PR 21-JUL-1999; 9905-0145086.
PR 21-JUL-1999; 9905-0145088.
PR 22-JUL-1999; 9905-0145085.
PR 22-JUL-1999; 9905-0145087.
PR 22-JUL-1999; 9905-0145089.
PR 22-JUL-1999; 9905-0145192.
PR 23-JUL-1999; 9905-0145145.
PR 23-JUL-1999; 9905-0145218.
PR 26-JUL-1999; 9905-0145224.
PR 27-JUL-1999; 9905-0145276.
PR 27-JUL-1999; 9905-0145913.
PR 27-JUL-1999; 9905-0145918.
PR 27-JUL-1999; 9905-0145919.
PR 28-JUL-1999; 9905-0145951.
PR 02-AUG-1999; 9905-0146386.
PR 02-AUG-1999; 9905-0146388.
PR 02-AUG-1999; 9905-0146389.
PR 03-AUG-1999; 9905-0147038.
PR 04-AUG-1999; 9905-0147203.
PR 04-AUG-1999; 9905-0147302.
PR 05-AUG-1999; 9905-0147192.
PR 05-AUG-1999; 9905-0147260.
PR 06-AUG-1999; 9905-0147303.
PR 06-AUG-1999; 9905-0147416.
PR 09-AUG-1999; 9905-0147493.
PR 09-AUG-1999; 9905-0147935.
PR 10-AUG-1999; 9905-0148171.
PR 11-AUG-1999; 9905-0148319.
PR 12-AUG-1999; 9905-0148341.
PR 13-AUG-1999; 9905-0148565.
PR 13-AUG-1999; 9905-0148684.
PR 16-AUG-1999; 9905-0149358.
PR 17-AUG-1999; 9905-0149175.
PR 18-AUG-1999; 9905-0149426.
PR 20-AUG-1999; 9905-0149722.
PR 20-AUG-1999; 9905-0149723.
PR 20-AUG-1999; 9905-0149929.
PR 23-AUG-1999; 9905-0149902.
PR 23-AUG-1999; 9905-0149930.
PR 25-AUG-1999; 9905-0150566.
PR 26-AUG-1999; 9905-0150884.
PR 27-AUG-1999; 9905-0151065.
PR 27-AUG-1999; 9905-0151066.
PR 27-AUG-1999; 9905-0151080.
PR 30-AUG-1999; 9905-0151303.
PR 31-AUG-1999; 9905-0151438.
PR 01-SEP-1999; 9905-0151930.
PR 07-SEP-1999; 9905-0152363.
PR 10-SEP-1999; 9905-0153070.

```
PR 13-SEP-1999; 990S-0153758.
PR 15-SEP-1999; 990S-0154018.
PR 16-SEP-1999; 990S-0154039.
PR 20-SEP-1999; 990S-0154779.
PR 22-SEP-1999; 990S-0155139.
PR 23-SEP-1999; 990S-0155486.
PR 24-SEP-1999; 990S-0155659.
PR 28-SEP-1999; 990S-0156458.
PR 29-SEP-1999; 990S-0156596.
PR 04-OCT-1999; 990S-0157117.
PR 05-OCT-1999; 990S-0157753.
PR 06-OCT-1999; 990S-0157865.
PR 07-OCT-1999; 990S-0158029.
PR 08-OCT-1999; 990S-0158332.
PR 12-OCT-1999; 990S-0158369.
PR 13-OCT-1999; 990S-0159293.
PR 13-OCT-1999; 990S-0159294.
PR 13-OCT-1999; 990S-0159295.
PR 14-OCT-1999; 990S-0159329.
PR 14-OCT-1999; 990S-0159330.
PR 14-OCT-1999; 990S-0159331.
PR 14-OCT-1999; 990S-0159637.
PR 14-OCT-1999; 990S-0159638.
PR 18-OCT-1999; 990S-0159584.
PR 21-OCT-1999; 990S-0160741.
PR 21-OCT-1999; 990S-0160767.
PR 21-OCT-1999; 990S-0160768.
PR 21-OCT-1999; 990S-0160770.
PR 21-OCT-1999; 990S-0160814.
PR 21-OCT-1999; 990S-0160815.
PR 22-OCT-1999; 990S-0160980.
PR 22-OCT-1999; 990S-0160981.
PR 22-OCT-1999; 990S-0160989.
PR 25-OCT-1999; 990S-0161404.
PR 25-OCT-1999; 990S-0161405.
PR 25-OCT-1999; 990S-0161406.
PR 26-OCT-1999; 990S-0161359.
PR 26-OCT-1999; 990S-0161360.
PR 26-OCT-1999; 990S-0161361.
PR 28-OCT-1999; 990S-0161920.
PR 28-OCT-1999; 990S-0161922.
PR 28-OCT-1999; 990S-0161993.
PR 29-OCT-1999; 990S-0162142.

alignment_scores:
      Quality: 226.50      length: 146
      Ratio: 2.157      Gaps: 4
Percent Similarity: 71.918      Percent Identity: 32.877

alignment_block:
US-09-676-718a-1 x AAC47500 ..

Align seg 1/1 to: AAC47500 from: 1 to: 682

17 LeuLeuLeuAlaThrValLeuGlnAlaValSerAlaPheGluAlaGluPh 33
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
41 TTGATATTGGCATCGACGATCTCAGCAAAAGAG.....CAGCT 78
33 eSerSerGluAlaCysArgGluLeuGluPheSerSerAsnLeuLeuGly 50
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
79 CAGACACTAAAGAAATGTAGAGATCTAGGGTTTAGGGCTTGACACTGCT 128
50 eSerSerCysAspLeuLeuGluGlnPhe...AsnLeuLeuGlnLeuAsp 65
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
129 CCGATTGCCACCTCAGCTCTCGAATACCTTAAGCAACCAAGTTGTAAT 178
66 AspCysArgGlyCysGlnGluGlnAlaGlnPheGluThrLysLysLe 82
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
179 GATTGCTTGAAGATGCTGCTGATGATTCGAGGATTCATGACCAAGGT 228
82 u...TyrAlaGluAlaLeuLeuGluValCysGly**LysLeuGlyArg 98
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
229 TACCTATTCTGGCGCTATATTATGAGGTGATGAGGAAGTGTGTTTCT 278
```

```
98 heProGlnValGlnAlaPheValArgSerAspLysProLysLeuPheArg 114
:||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
279 ATCCGAAATTTGTTGTTTCATTGAAAGAGAAAGAAAG...TTCCCT 325
115 GlyLeuGlnIleLysTyrValArgGlySerAspProValLeuLysLeu 131
:||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
326 AGCGTCAAGCTTCAGTACATTTCACCTCACCCACCAAGTTGATCATGCT 375
131 uAspAspAsnGlyAsnIleLeaGluGluLeuSerIleLeuLysTyrPasn 148
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
376 CGACGAAGACGGTCAACATAAGCAATCATATTGATGACAACTGGAAAC 425
148 hrAspSerValGluGluPheLeuSerGluLysLeuGlu 160
:::|::|:|||||:|||||:|||||:|||||:|||||:|||||:|||||
426 GCGAACAATCTACTCAGTATATGCGAGAGAAGTCAAG 463

seq_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAC34325
seq_documentation_block:
ID AAC34325 standard; DNA; 695 BP.
XX
AC AAC34325;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 6246.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 990S-0121825.
PR 05-MAR-1999; 990S-0123180.
PR 09-MAR-1999; 990S-0123548.
PR 23-MAR-1999; 990S-0125788.
PR 25-MAR-1999; 990S-0126264.
PR 29-MAR-1999; 990S-0126785.
PR 01-APR-1999; 990S-0127462.
PR 06-APR-1999; 990S-0128234.
PR 08-APR-1999; 990S-0128714.
PR 16-APR-1999; 990S-0129845.
PR 19-APR-1999; 990S-0130077.
PR 21-APR-1999; 990S-0130449.
PR 23-APR-1999; 990S-0130510.
PR 28-APR-1999; 990S-0131449.
PR 30-APR-1999; 990S-0132048.
PR 04-MAY-1999; 990S-0132407.
PR 05-MAY-1999; 990S-0132484.
PR 06-MAY-1999; 990S-0132485.
PR 06-MAY-1999; 990S-0132486.
PR 06-MAY-1999; 990S-0132487.
PR 07-MAY-1999; 990S-0132863.
PR 11-MAY-1999; 990S-0134256.
PR 14-MAY-1999; 990S-0134218.
PR 14-MAY-1999; 990S-0134219.
PR 14-MAY-1999; 990S-0134221.
PR 14-MAY-1999; 990S-0134370.
PR 18-MAY-1999; 990S-0134768.
PR 19-MAY-1999; 990S-0134941.
PR 20-MAY-1999; 990S-0135124.
PR 21-MAY-1999; 990S-0135353.
PR 24-MAY-1999; 990S-0135629.
PR 25-MAY-1999; 990S-0136021.
```

PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139753.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139839.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145226.
PR 26-JUL-1999; 99US-0145591.
PR 27-JUL-1999; 99US-0145591.
PR 27-JUL-1999; 99US-0145591.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.

PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.


```

PR 29-OCT-1999; 99US-0162142.
alignment_scores:
    Quality: 226.50      Length: 146
    Ratio: 2.157        Gaps: 4
Percent Similarity: 71.918    Percent Identity: 32.877

alignment_block:
US-09-676-718A-1 x AAC34325 ..

Align seg 1/1 to: AAC34325 from: 1 to: 695

17 LeuLeuLeuAlaThrValLeuGlnAlaValSerAlaPheGlyAlaGluPh 33
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
54 TTGATGTTGGCATCGAGATCTCGACAAAGAG.....CAGCT 91
33 eSerSerGluAlaCysArgGluLeuGlyPheSerSerAsnLeuLeuGly 50
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
92 CAGCACAAGAAATGTGAGCATCTAGGGTTTACGGCTTCTGCTCTGCT 141
50 eSerCysAspLeuLeuGlyGlnPhe..AsnLeuLeuGlnLeuAspPro 65
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
142 CCGATTCACACTACTCTCGAATACGTTAAGACCAAGAGTTGTAATCT 191
66 AspCysArgGlyCysCysGlnGluAlaGlnPheGluThrLysLysLe 82
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
192 GATTCGTTGAAATGTTGTGCTGATGATTCGAGGATTCATGACCAAGT 241
82 u...TyrAlaGlyAlaIleLeuGlnValCysGly**LysLeuGlyLArg 98
   : |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
242 TACCTATTCGCGCCTATATTAGAGTGTGTATGAGAAAGTTGGTTTCT 291
98 heProGlnValGlnAlaPheValArgSerAspLysProLysLeuPheArg 114
   : |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
292 ATCTCGAAATGTTGTGTTTCATTGACAGAGAAAGAAAG...TTCCT 338
115 GlyLeuGlnIleLysTyrValArgGlySerAspProValLeuLysLeu 131
   : |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
339 AGCGTCAAAAGTTCAGTACATTTTCACACTCACCCCAAGTTGATCATGCT 388
131 uAspAspAspGlyAsnIleAlaGluGlnLeuSerIleLeuLysTrpAsnT 148
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
389 CGACGAAAGACGCTGACATAGCAATCAATTAGATGCACACTGCAAAAC 438
148 hrAspSerValGluGluPheLeuSerGluLysLeuGln 160
   : |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
439 GCGAACAATCTACTGCACTATATGCGAGAGAAAGTCAAG 476

seq_name: /SIDSL/gcgsdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:ABA57362

seq_documentation_block:
ID ABA57362 standard; DNA; 483 BP.
XX
AC ABA57362;
XX
DT 01-FEB-2002 (first entry)
XX
DE Human foetal liver single exon nucleic acid probe #5667.
XX
KW Human, foetal liver, gene expression, single exon nucleic acid probe, ss.
XX
OS Homo sapiens.
XX
PN WO200157277-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00669.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.

```

```

PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483447/52.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
XX
PS Claim 1; SEQ ID NO 5667; 639bp + sequence listing; English.
XX
CC The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 483 BP; 139 A; 91 C; 105 G; 148 T; 0 other;

alignment_scores:
    Quality: 172.50      Length: 86
    Ratio: 2.875        Gaps: 1
Percent Similarity: 69.767    Percent Identity: 51.163

alignment_block:
US-09-676-718A-1 x ABA57362 ..

Align seg 1/1 to: ABA57362 from: 1 to: 483

1 MetaIaIaGlyProSerGlyCysLeuValProIaPheGlyLysArgLe 17
   : |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
241 GTGCGACAGAGAGAGATGATGCTAGACTGCATTTGGAGTCAAGT 290
17 uLeuLeuAlaThrValLeuGlnAlaValSerAlaPheGlyAlaGluPhes 34
   : |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
291 GCCAAGAGCAACTGATCCCAAGCGCTAGCTGATTTGACAGACTGCTAC 340
34 eSerSerGluAlaCysArgGluLeuGlyPheSerSerAsnLeuLeuLysSer 50
   : |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
341 TTCTCTT.....GGCTCTCTTCACCAACTTCTCTCTGTTGC 374
51 SerCysAspLeuLeuGlyGlnPheAsnLeuLeuGlnLeuAspProAspCy 67
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
375 TCTTGTGACCTTCTGACACTGTTTACCTGTTCAGCTGGATCTTAAGT 424
67 sArgGlyCysCysGlnGluGlnAlaGlnPheGluThrLysLysLeuTyrA 84
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
425 CAGAGACTACTATCAGAGAAAGTTCAATTGAAACAGAAAGCTCGATG 474
84 IaglyAla 86
   : |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
475 TAGGAGCT 482

seq_name: /SIDSL/gcgsdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAK05398

seq_documentation_block:
ID AAK05398 standard; DNA; 483 BP.
XX
AC AAK05398;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe SEQ ID NO: 5389.
XX

```

KW Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer; ss.

OS Homo sapiens.

PN W0200157275-A2.

PD 09-AUG-2001.

30-JAN-2001; 2001WO-US00667.

PR 04-FEB-2000; 2000US-0180312.

PR 30-JUN-2000; 2000US-0608408.

PR 21-SEP-2000; 2000US-0234687.

PR 04-OCT-2000; 2000GB-0024263.

PA (MOLE-) MOLECULAR DYNAMICS I

PI Penn SG, Hanzel DK, Chen W

DR WPI; 2001-483446/52.

PT Single exon nucleic acid proc

XX

33 XX

CC probes which are derived from

cc which may enable the diagnosis

CC epilepsy and cancers. The pr

0
C
67
A.
130
ED.
493
C
C
C
C
C
C
C
XX
XX

alignment_scores: 177 50

Ratio: 2.875

0
1
2
3
4
5
6
7
8
9

US-09-676-718A-1 x AAK05398 ..

Align seg 1/1 to: AAK05398 from

1 MetaLaalaglyProserGlyCys

241 GTGGCAGCAGGAGAGAGTACATGTT

17 uleuAlaThrValLeuGlnAla

291 GCCAAGAGCACTGCATCCCAAGC

34 erSerGluAlaCysArgGluLeu

341 TTTCTTT.....G

51 sercysaspleuengllycInphe

3/5 TCTTGTGACCCCTTCTGGACCTGT

b7c
b7d
b7e
b7f
b7g
b7h
b7i
b7j
b7k
b7l
b7m
b7n
b7o
b7p
b7q
b7r
b7s
b7t
b7u
b7v
b7w
b7x
b7y
b7z
b7aa
b7ab
b7ac
b7ad
b7ae
b7af
b7ag
b7ah
b7ai
b7aj
b7ak
b7al
b7am
b7an
b7ao
b7ap
b7aq
b7ar
b7as
b7at
b7au
b7av
b7aw
b7ax
b7ay
b7az
b7ba
b7bb
b7bc
b7bd
b7be
b7bf
b7bg
b7bh
b7bi
b7bj
b7bk
b7bl
b7bm
b7bn
b7bo
b7bp
b7bq
b7br
b7bs
b7bt
b7bu
b7bv
b7bw
b7bx
b7by
b7bz
b7ca
b7cb
b7cc
b7cd
b7ce
b7cf
b7cg
b7ch
b7ci
b7cj
b7ck
b7cl
b7cm
b7cn
b7co
b7cp
b7cq
b7cr
b7cs
b7ct
b7cu
b7cv
b7cw
b7cx
b7cy
b7cz
b7da
b7db
b7dc
b7dd
b7de
b7df
b7dg
b7dh
b7di
b7dj
b7dk
b7dl
b7dm
b7dn
b7do
b7dp
b7dq
b7dr
b7ds
b7dt
b7du
b7dv
b7dw
b7dx
b7dy
b7dz
b7ea
b7eb
b7ec
b7ed
b7ee
b7ef
b7eg
b7eh
b7ei
b7ej
b7ek
b7el
b7em
b7en
b7eo
b7ep
b7eq
b7er
b7es
b7et
b7eu
b7ev
b7ew
b7ex
b7ey
b7ez
b7fa
b7fb
b7fc
b7fd
b7fe
b7ff
b7fg
b7fh
b7fi
b7fj
b7fk
b7fl
b7fm
b7fn
b7fo
b7fp
b7fq
b7fr
b7fs
b7ft
b7fu
b7fv
b7fw
b7fx
b7fy
b7fz
b7ga
b7gb
b7gc
b7gd
b7ge
b7gf
b7gg
b7gh
b7gi
b7gj
b7gk
b7gl
b7gm
b7gn
b7go
b7gp
b7gq
b7gr
b7gs
b7gt
b7gu
b7gv
b7gw
b7gx
b7gy
b7gz
b7ha
b7hb
b7hc
b7hd
b7he
b7hf
b7hg
b7hh
b7hi
b7hj
b7hk
b7hl
b7hm
b7hn
b7ho
b7hp
b7hq
b7hr
b7hs
b7ht
b7hu
b7hv
b7hw
b7hx
b7hy
b7hz
b7ia
b7ib
b7ic
b7id
b7ie
b7if
b7ig
b7ih
b7ii
b7ij
b7ik
b7il
b7im
b7in
b7io
b7ip
b7iq
b7ir
b7is
b7it
b7iu
b7iv
b7iw
b7ix
b7iy
b7iz
b7ja
b7jb
b7jc
b7jd
b7je
b7jf
b7jg
b7jh
b7ji
b7jj
b7jk
b7jl
b7jm
b7jn
b7jo
b7jp
b7jq
b7jr
b7js
b7jt
b7ju
b7jv
b7jw
b7jx
b7jy
b7jz
b7ka
b7kb
b7kc
b7kd
b7ke
b7kf
b7kg
b7kh
b7ki
b7kj
b7kk
b7kl
b7km
b7kn
b7ko
b7kp
b7kq
b7kr
b7ks
b7kt
b7ku
b7kv
b7kw
b7kx
b7ky
b7kz
b7la
b7lb
b7lc
b7ld
b7le
b7lf
b7lg
b7lh
b7li
b7lj
b7lk
b7ll
b7lm
b7ln
b7lo
b7lp
b7lq
b7lr
b7ls
b7lt
b7lu
b7lv
b7lw
b7lx
b7ly
b7lz
b7ma
b7mb
b7mc
b7md
b7me
b7mf
b7mg
b7mh
b7mi
b7mj
b7mk
b7ml
b7mm
b7mn
b7mo
b7mp
b7mq
b7mr
b7ms
b7mt
b7mu
b7mv
b7mw
b7mx
b7my
b7mz
b7na
b7nb
b7nc
b7nd
b7ne
b7nf
b7ng
b7nh
b7ni
b7nj
b7nk
b7nl
b7nm
b7nn
b7no
b7np
b7nq
b7nr
b7ns
b7nt
b7nu
b7nv
b7nw
b7nx
b7ny
b7nz
b7oa
b7ob
b7oc
b7od
b7oe
b7of
b7og
b7oh
b7oi
b7oj
b7ok
b7ol
b7om
b7on
b7oo
b7op
b7oq
b7or
b7os
b7ot
b7ou
b7ov
b7ow
b7ox
b7oy
b7oz
b7pa
b7pb
b7pc
b7pd
b7pe
b7pf
b7pg
b7ph
b7pi
b7pj
b7pk
b7pl
b7pm
b7pn
b7po
b7pp
b7pq
b7pr
b7ps
b7pt
b7pu
b7pv
b7pw
b7px
b7py
b7pz
b7qa
b7qb
b7qc
b7qd
b7qe
b7qf
b7qg
b7qh
b7qi
b7qj
b7qk
b7ql
b7qm
b7qn
b7qo
b7qp
b7qq
b7qr
b7qs
b7qt
b7qu
b7qv
b7qw
b7qx
b7qy
b7qz
b7ra
b7rb
b7rc
b7rd
b7re
b7rf
b7rg
b7rh
b7ri
b7rj
b7rk
b7rl
b7rm
b7rn
b7ro
b7rp
b7rq
b7rr
b7rs
b7rt
b7ru
b7rv
b7rw
b7rx
b7ry
b7rz
b7sa
b7sb
b7sc
b7sd
b7se
b7sf
b7sg
b7sh
b7si
b7sj
b7sk
b7sl
b7sm
b7sn
b7so
b7sp
b7sq
b7sr
b7ss
b7st
b7su
b7sv
b7sw
b7sx
b7sy
b7sz
b7ta
b7tb
b7tc
b7td
b7te
b7tf
b7tg
b7th
b7ti
b7tj
b7tk
b7tl
b7tm
b7tn
b7to
b7tp
b7tq
b7tr
b7ts
b7tt
b7tu
b7tv
b7tw
b7tx
b7ty
b7tz
b7ua
b7ub
b7uc
b7ud
b7ue
b7uf
b7ug
b7uh
b7ui
b7uj
b7uk
b7ul
b7um
b7un
b7uo
b7up
b7uq
b7ur
b7us
b7ut
b7uu
b7uv
b7uw
b7ux
b7uy
b7uz
b7va
b7vb
b7vc
b7vd
b7ve
b7vf
b7vg
b7vh
b7vi
b7vj
b7vk
b7vl
b7vm
b7vn
b7vo
b7vp
b7vq
b7vr
b7vs
b7vt
b7vu
b7vv
b7vw
b7vx
b7vy
b7vz
b7wa
b7wb
b7wc
b7wd
b7we
b7wf
b7wg
b7wh
b7wi
b7wj
b7wk
b7wl
b7wm
b7wn
b7wo
b7wp
b7wq
b7wr
b7ws
b7wt
b7wu
b7wv
b7ww
b7wx
b7wy
b7wz
b7xa
b7xb
b7xc
b7xd
b7xe
b7xf
b7xg
b7xh
b7xi
b7xj
b7xk
b7xl
b7xm
b7xn
b7xo
b7xp
b7xq
b7xr
b7xs
b7xt
b7xu
b7xv
b7xw
b7xx
b7xy
b7xz
b7ya
b7yb
b7yc
b7yd
b7ye
b7yf
b7yg
b7yh
b7yi
b7yj
b7yk
b7yl
b7ym
b7yn
b7yo
b7yp
b7yq
b7yr
b7ys
b7yt
b7yu
b7yv
b7yw
b7yx
b7yy
b7yz
b7za
b7zb
b7zc
b7zd
b7ze
b7zf
b7zg
b7zh
b7zi
b7zj
b7zk
b7zl
b7zm
b7zn
b7zo
b7zp
b7zq
b7zr
b7zs
b7zt
b7zu
b7zv
b7zw
b7zx
b7zy
b7zz
b7aa
b7ab
b7ac
b7ad
b7ae
b7af
b7ag
b7ah
b7ai
b7aj
b7ak
b7al
b7am
b7an
b7ao
b7ap
b7aq
b7ar
b7as
b7at
b7au
b7av
b7aw
b7ax
b7ay
b7az
b7ba
b7bb
b7bc
b7bd
b7be
b7bf
b7bg
b7bh
b7bi
b7bj
b7bk
b7bl
b7bm
b7bn
b7bo
b7bp
b7bq
b7br
b7bs
b7bt
b7bu
b7bv
b7bw
b7bx
b7by
b7bz
b7ca
b7cb
b7cc
b7cd
b7ce
b7cf
b7cg
b7ch
b7ci
b7cj
b7ck
b7cl
b7cm
b7cn
b7co
b7cp
b7cq
b7cr
b7cs
b7ct
b7cu
b7cv
b7cw
b7cx
b7cy
b7cz
b7da
b7db
b7dc
b7dd
b7de
b7df
b7dg
b7dh
b7di
b7dj
b7dk
b7dl
b7dm
b7dn
b7do
b7dp
b7dq
b7dr
b7ds
b7dt
b7du
b7dv
b7dw
b7dx
b7dy
b7dz
b7ea
b7eb
b7ec
b7ed
b7ee
b7ef

425 CAGAGAGTACTATCAGAGGAAAGT

84 JAGLYALA 86

•
•
•
•

475 TAGGAGCT 482

OM of: US-09-676-718a-1 to: GenEmbl:* out_format : pfs
Date: Sep 18, 2002 5:39 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-MODEL=firmer-p2n.model -DEV=xlh  
-Q/cg2.1/USPTO.spool/US09676718/r/unat_17092002.143433.21305/app-query.fasta_1.222  
-DB=GenEmbl -QFMT=fastap -SUFFIX=ge -GAPOP=12.000 -GAPEXT=4.000  
-NMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -OGAPOP=4.500  
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000  
-FGAPEXT=7.000 -TGAPOP=10.000 -YGAPEXT=0.500 -DELDP=6.000  
-DELEXT=7.000 -START=1 -MATRIX=blsmb2 -TRANS=human40.cd1  
-LIST=45 -DOCALL=200 -THR_SCORE=pcr -THR_MAX=100 -THR_MIN=0  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZE=500  
-MINLEN=0 -MAXLEN=200000000 -USER=US09676718.acgcn1.1.3896  
-NCPU=6 -ICPU=3 -LONGSDOS -DEV_TIMEOUT=120 -WARN_TIMEOUT=30  
-NO_XLPEXT -WAIT -THREADS=1
```

Search information block:

Query: US-09-676-718a-1
Query length: 162
Database: GenEmbl:*
Database sequences: 1797656
Database length: 187333701
Search time (sec): 1928.670000

score_list:

Sequence	Strd Orig	ZScore	EScore	len	Documentation
gb_pr:AF051894	+	819.00	1711.94	5.4e-87	1244
gb_pr:BC005294	+	819.00	1711.55	5.7e-87	1299
gb_pr:AF288991	+	819.00	1710.13	6.8e-87	1519
gb_pr:BC016359	+	819.00	1709.88	7.1e-87	1562
gb_to:AF288740	+	747.00	1558.10	2.0e-78	1463
gb_to:BC019792	+	745.00	1555.38	2.8e-78	1238
gb_to:AF390544	+	741.00	1546.93	8.4e-78	1236
gb_to:BC021697	+	701.50	1463.24	3.9e-73	1249
gb_to:BC010481	+	476.00	987.86	1.2e-46	1020
gb_pat:AX330345	+	392.00	820.81	2.3e-37	313
gb_in:AF457547	+	383.00	793.50	6.0e-36	624
gb_pr:AF288992	+	304.00	586.28	2.7e-24	64165
gb_hlg:AL335981	+	304.00	578.23	7.6e-24	155871
gb_pr:AF267982S2	+	303.50	617.94	4.7e-26	1743
gb_pr:HS050445	+	303.50	579.02	6.9e-24	127201
gb_hlg:AC096112	+	284.00	533.34	1.9e-21	165889
gb_hlg:AC019594	+	274.00	521.75	1.1e-20	71971
gb_in:AC010054	+	274.00	513.79	3.0e-20	173108
gb_in:AC009375	+	274.00	513.40	3.1e-20	180787
gb_in:AE003523	+	274.00	509.32	5.3e-20	283337
gb_pr:AF267982S5	+	217.00	433.78	8.5e-16	1968
gb_hlg:AL335981	+	215.00	394.11	1.4e-13	155871
gb_pr:AC004925	+	159.00	336.99	2.0e-13	154959
gb_sts:G25431	+	159.00	332.68	3.6e-10	181
gb_sts:G25431	+	159.00	332.68	3.6e-10	181
gb_pl:AC007153	+	141.00	237.01	7.7e-05	103223
gb_pr:HS74R3F	+	125.00	256.42	6.4e-06	291
gb_sts:HSPE29R10	+	125.00	253.28	9.6e-06	411
gb_pr:AF267982S1	+	123.50	196.76	6.0e-05	147146
gb_hlg:AP003316	+	123.50	196.76	6.0e-05	147146
gb_hlg:AP003316	+	123.50	196.76	6.0e-05	147146
gb_pr:HS74R3R	+	109.00	172.01	0.3225	76485
gb_in:AC024211	+	109.00	172.01	0.3225	76485
gb_hlg:AC006907	+	104.50	154.28	3.13	188972
gb_hlg:AC006112	+	104.00	154.41	3.08	165889
gb_pr:AF267982S3	+	100.00	159.99	0.0283	1157
gb_to:AV043488	+	92.50	179.84	0.1181	687
gb_to:BC019742	+	92.50	180.26	0.1119	584
gb_pr:HS1068E13	+	91.00	127.92	92.14	148177

gb_hlg:AC093424 + 88.50 122.29 189.73 133853 | AC093424 Homo sapiens chr
gb_hlg:AC093199 - 88.50 121.82 201.40 161951 | AC093199 Rattus norvegicus
gb_hlg:AC025793 - 88.50 121.61 206.84 165704 | AC025793 Homo sapiens chr
em_hlg:inv:AC080697 + 85.50 161.46 1.25 1018 | AC080697 Giardia intestinal
gb_pr:AF267982S4 + 85.00 162.79 1.05 783 | AF267985 Homo sapiens 15 kda

seq_name: gb_pr:AF051894

seq_documentation_block:

LOCUS AF051894 1244 bp mRNA linear PRI 26-OCT-2000
DEFINITION Homo sapiens 15 kda selenoprotein mRNA, complete cds.
ACCESSION AF051894
VERSION AF051894.1 GI:3095110

KEYWORDS

SOURCE

ORGANISM

human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

AUTHORS

TITLE

JOURNML

FEATURES

SOURCE

CDS

misc-feature

BASE COUNT

ORIGIN

alignment_scores:

Percent Similarity:

Percent Identity:

alignment_block:

US-09-676-718a-1 x AF051894

Align seg 1/1 to: AF051894 from: 1 to: 1244

1 MetAlaAlaGlyProSerGlyCysLeuValProAlaPheGlyIysArgLe 17

5 ATGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 54

17 uleuLeuAlaThrValLeuGlnAlaValSerAlaPheGlyAlaGlyPhe 34

55 GTTGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 104

34 erSerGluAlaCysArgGluLeuGlyPheSerSerAsnLeuLeuCysSer 50

```

|||||
105 CATGGAGGCGTGCAGAGAGTTAGCTTTTCTACCAACTTCTTGACGC 154
51 SerCysAspLeuLeuGlyGlnPheAsnLeuLeuGlnLeuAspProAspCy 67
|||||
155 TCTTGATGATCTCTCGACAGATTCAACCTGCTTGATGATGATG 204
67 sArgGlyCysCysGlnGlnGlnPheGlnPheGlnPheGlnPheGlnPhe 84
|||||
205 CAGGAGATGCTGTCAGAGAGCAATTTGAACCAAAAAGCTGATG 254
84 laGlyAlaIleLeuGluValCysGly**LysLeuGlyArgPheProGln 100
|||||
255 CAGAGACTATCTTGAAAGTTGTGATGAAATTTGGAGAGTTCCCTCAA 304
101 ValGlnAlaPheValArgSerAspLysProLysLeuPheArgGlyLeuG 117
|||||
305 GTCCAACTTTTGTAGAGATGATTAACCCAAACTGTTCAAGAGACGCA 354
117 nIleLysTyrValArgGlySerAspProValLeuLysLeuAspAspA 134
|||||
355 AATCAAGTATGCTCGTGCAGACCTGTATTAAGCTTTTGACGACGA 404
134 snGlyAsnIleAlaGlnGlnLeuSerIleLeuLysTyrPasnThrAspSer 150
|||||
405 ATGGGAACATTGCTGAAGAACTGAGCATTTCTCAATGGAACACAGACAGT 454
151 ValGlnGlnPheLeuSerGlnLysLeuGlnArgGlyLeu 162
|||||
455 GTAGAGAAATTCCTGAGTGAAGAAAGTTGGAACGACATA 490
seq_name: gb_pr:BC005294
seq_documentation_block:
LOCUS BC005294 1299 bp mRNA linear PRI 12-JUL-2001
DEFINITION Homo sapiens, clone MGC:12358 IMAGE:4041321, mRNA, complete cds.
ACCESSION BC005294
VERSION BC005294.1 GI:13529007
KEYWORDS MGC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1299)
Strausberg, R.
Direct Submission
Submitted (27-MAR-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL).
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www-shgc.stanford.edu
Contact: (Dickson, Mark) mcd@paxll.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov
Series: IRAL Plate: 16 Row: b Column: 12
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 3095110.
Location/Qualifiers
1..1299
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:12358 IMAGE:4041321"

```

```

/tissue_type="Kidney, hypernephroma"
/clone_id="NIH_MGC_58"
/lab_host="DH10B"
/notes="Vector: pDNR-LIB"
31..309
/codon_start=1
/product="Unknown (protein for MGC:12358)"
/protein_id="AAH05294.1"
/db_xref="GI:13529008"
/translation="MAGPSGCLVPAPFGRLLRLATLVQAVSAFGAEFSSEACRELGFS
SNLCSSCDLQGFNLIDLPDRCGCCOEBAQFETIKLYAGAILVEVC"
BASE COUNT 419 a 223 c 264 g 393 t
ORIGIN
alignment_scores:
Quality: 819.00 Length: 162
Ratio: 5.119 Gaps: 0
Percent Similarity: 98.765 Percent Identity: 98.765.
alignment_block:
US-09-676-718A-1 x BC005294 ..
Align seg 1/1 to: BC005294 from: 1 to: 1299
1 MetAlaAlaGlyProSerGlyCysLeuValProAlaPheGlyLysArgLe 17
|||||
31 ATGCGCGCTGGCGCCAGTGGCTGTGTCGCCGCCGCTTGCGGCGTACGGTT 80
17 uLeuLeuAlaThrValLeuGlnAlaValSerAlaPheGlyValAlaLubPhe 34
|||||
81 GTTGTGGCGACTGTGCTTCAAGCGGCTGTCTGCTTTGGGGCAGAGTTT 130
34 eSerGlnAlaCysArgGlnuLeuGlyPheSerSerAsnLeuLeuCysSer 50
|||||
131 CATGGAGGCGATGCAGAGAGATTAGCTTTTCTACCAACTTCTTGACGC 180
51 SerCysAspLeuLeuGlyGlnPheAsnLeuLeuGlnLeuAspProAspCy 67
|||||
181 TCTTGATGATCTCTCGACAGATTCAACCTGCTTCAGCTGATGATG 230
67 sArgGlyCysCysGlnGlnGlnPheGlnPheGlnPheGlnPheGlnPhe 84
|||||
231 CAGAGATGCTGTCAGAGAGAAACCAATTTGAAACCAAAAAGCTGATG 280
84 laGlyAlaIleLeuGluValCysGly**LysLeuGlyArgPheProGln 100
|||||
281 CAGAGGCTATCTTGAAAGTTGTGAGTGAAGAAATTTGGGAAGTTCCCTCAA 330
101 ValGlnAlaPheValArgSerAspLysProLysLeuPheArgGlyLeuG 117
|||||
331 GTCCAACTTTTGTAGAGATGATTAACCCAAACTGTTCAAGAGACGCA 380
117 nIleLysTyrValArgGlySerAspProValLeuLysLeuAspAspA 134
|||||
381 AATCAAGTATCTCGTGGTTCAGACCTGTATTAAAGCTTTTGAACGCA 430
134 snGlyAsnIleAlaGlnGlnLeuSerIleLeuLysTyrPasnThrAspSer 150
|||||
431 ATGGGAACATTGCTGAAGAACTGAGCATTTCTCAATGGAACACAGACAGT 480
151 ValGlnGlnPheLeuSerGlnLysLeuGlnArgGlyLeu 162
|||||
481 GTAGAGAAATTCCTGAGTGAAGAAAGTTGGAACGACATA 516
seq_name: gb_pr:AF288991
seq_documentation_block:
LOCUS AF288991 1519 bp mRNA linear PRI 12-NOV-2000
DEFINITION Homo sapiens 15 kDa selenoprotein (SEP15) mRNA, complete cds;
alternatively spliced.
ACCESSION AF288991
VERSION AF288991.1 GI:11138954

```

KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1519)
AUTHORS Kumaraswamy, E., Malikh, A., Korotkov, K.V., Kozayvkin, S., Hu, Y., Kwon, S.Y., Moustafa, M.E., Carlson, B.A., Berry, M.J., Lee, B.J., Hatfield, D.L., Diamond, A.M. and Gladyshev, V.N.
TITLE Structure-expression relationships of the 15 kDa selenoprotein gene: possible role of the protein in cancer etiology
JOURNAL J Biol Chem. 275 (45), 35540-35547 (2000)
PUBMED 10945981
REFERENCE 2 (bases 1 to 1519)
AUTHORS Kumaraswamy, E., Malikh, A., Korotkov, K.V., Kozayvkin, S., Hu, Y., Kwon, S., Moustafa, M.E., Carlson, B.A., Berry, M.J., Lee, B.J., Hatfield, D.L., Diamond, A.M. and Gladyshev, V.N.
TITLE Direct Submission
JOURNAL Submitted (21-JUL-2000) Department of Biochemistry, University of Nebraska, N151 Beadle Center, Lincoln, NE 68588, USA
FEATURES
source location/Qualifiers
1..1519
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="1p31"
1..1519
/gene="SEP15"
5..493
/gene="SEP15"
/note="alternatively spliced; long alternative transcript; contains SECIS element; selenocysteine"
/transl_except=(pos:281..283,aa:OTHER)
/codon_start=1
/product="15 kDa selenoprotein"
/protein_id="AA031556.1"
/db_xref="GI:11138955"
/translation="MAAGPSGCLVPAGFLRLILATVLAQVAFGEFSSSEACREIGFS
SNLSSCDLLOFNLLQIDPDRCGCCOEEQFETKILYAGAILLEVCCKLGRFOVQ
AFVRSRDKRLFRGLQIKYVRGSDPVLKLLDNGNTAEELSLIKMNTDSVEPTLSKLE
RI"
BASE COUNT 458 a 245 c 292 g 524 t
ORIGIN
alignment_scores:
Quality: 819.00 Length: 162
Ratio: 5.119 Gaps: 0
Percent Similarity: 98.765 Percent Identity: 98.765
alignment_block:
US-09-676-718A-1 x AF288991 ..
Align seg 1/1 to: AF288991 from: 1 to: 1519
1 MetAlAlAGlYProSerGlyCysLeuValProAlAPheGlyLysArgLe 17
|||||
5 ATGGCGGGTGGCGCCAGTGGGTCTGTGTCGCCGGCTTGGCTACGGTT 54
|||||
17 uleuLeuAlAthrValLeuGlnAlaValSerAlAPheGlyAlAGluPheS 34
|||||
55 GTTGTGGGGCAGCTGCTTCACGCGGTCTGTCTTTGGGGCAGAGTATT 104
|||||
34 exSerGluAlaCysArgGluLeuGlyPheSerSerAsnLeuLeuCysSer 50
|||||
105 CATCGAGGAGCATGCAGAGACTTTCAGCAACTTCTTTCGACG 154
|||||
51 SerCysAspLeuLeuGlnPheAsnLeuLeuGlnLeuAspProAspCy 67
|||||
155 TCTTGTGATCTTCTCGACAGTTCACACTGCTTCAGCTGATCCTGATG 204
|||||
67 sArgGlyCysCysGlnGluGlnAlaGlnPheGlyLysLysLeuLeuTyr 84
|||||

205 CAGAGCATGCTGTACAGAGACACACATTTGAACCAAAAGCTGTATG 254
84 laGlyAlaIleLeuGluValCysGly**LysLeuGlyArgPheProGln 100
|||||
255 CAGAGCATGCTGTGAAGTTTGTGGATGAAATTTGGAAAGTTCCTCCAA 304
|||||
101 ValGlnAlAPheValArgSerAspLysProLysLeuPheArgGlyLeuG 117
|||||
305 GTCCAGCTTTTGTAGAGTGTAAACCCAACTGTTCAGAGACTGCA 354
|||||
117 nileLysTyrValArgLysSerAspProValLeuLysLeuLeuAspAsp 134
|||||
355 AATCAAGTATGCTCCGTGTCACACCTGTATTAACTTTTGGACGACA 404
|||||
134 snGlyAsnIleAlaGluGluLeuSerIleLeuLysTyrAsnThrAspSer 150
|||||
405 ATGGGAACATGCTGGAAGAACTGAGCTTCTCAATGGAAACACAGACGT 454
|||||
151 ValGluGluPheLeuSerGluLysLeuGluArgIle 162
|||||
455 GTAGAAAGAAATTCCTGAGTGAAGATTGGAACGCATA 490
seq_name: gb_pr:BC016359
seq_documentation_block:
LOCUS BC016359 1562 bp mRNA linear PRI 05-NOV-2001
DEFINITION Homo sapiens, clone MGC:24512 IMAGE:4096759, mRNA, complete cds.
ACCESSION BC016359
VERSION BC016359.1 GI:16741016
KEYWORDS MGC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1562)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (29-OCT-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
COMMENT Contact: MGC help desk
Email: cga@bs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbioology.org>
Contact: amadane@systemsbiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL Plate: 32 Row: 9 Column: 15
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis.
FEATURES
source location/Qualifiers
1..1562
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:24512 IMAGE:4096759"
/tissue_type="Bone marrow, chronic myelogenous leukemia"
/clone_id="N1H_MGC_54"
/lab_host="DH10B"
/note="vector: pDNR-LIB"
18..296
/codon_start=1
/product="Unknown (protein for MGC:24512)"
/protein_id="AAH16359.1"
/db_xref="GI:16741017"

```

/translation="MAGPSGCVLPARGILRLILATYQANVARGAERSRACRELGFS
SNLCSGCDLGGFNLPIDPDRGCCQEPQAEFTKLYAGAILVEGCG"
BASE COUNT      492 a      246 c      298 g      526 t
ORIGIN

alignment_scores:
    Quality: 819.00      Length: 162
    Ratio: 5.119      Gaps: 0
    Percent Similarity: 98.765      Percent Identity: 98.765

alignment_block:
US-09-676-718A-1 x BC016359

Align seg 1/1 to: BC016359 from: 1 to: 1562

1  MetaAlaIaGlyProSerGlyCysLeuValProAlaPheGlyLysArgLe 17
|||||
18  ATGGCGGCGTGGCCGAGTGGGTCTGTGGTGGCGGCTTTGGGCTACGGTT 67
|||||
17  uLeuLeuAlaThrValLeuGlnAlaValSerAlaPheGlyAlaGluPheS 34
|||||
68  GTTGTGGCGACTGTGCTTCAAGCGGTCTGCTTTGGGGCAGAGTTT 117
|||||
34  eSerGluAlaCysArgGluLeuGlyPheSerSerAsnLeuLeuCysSer 50
|||||
118  CATCGAGAGCATGACAGAGATTAGGCTTTTCTAGCAACTGCTTGCAGC 167
|||||
51  SerCysAspLeuLeuGlyGlnPheAsnLeuGlnLeuAspProAspCys 67
|||||
168  TCTTGTGATCTTCTCGACAGTTCACCTGCTTGCAGCTGATCCTGATTG 217
|||||
67  sArgGlyCysCysGlnGluGlnAlaGlnPheGluThrLysLysLeuTYR 84
|||||
218  CAGAGAGATGCTGTGAGGAGGAGCAAAATTGAAACCAAAAGCTGATAG 267
|||||
84  IaGlyAlaIleLeuGlnValCysGly**LysLeuGlyArgPheProGln 100
|||||
268  CAGGAGCTATTCTTGAAGTTGTGATGAAATGGGAAGTTCCTCA 317
|||||
101  ValGlnAlaPheValArgSerAspLysProLysLeuPheArgGlyLeuG1 117
|||||
318  GTCCAAAGCTTTTGTAGAGATGATAAACCAACTGTCAGAGCACTGCA 367
|||||
117  nIleLysTYRValArgGlySerAspProValLeuLysLeuLeuAspAspA 134
|||||
368  AATCAAGATGTCCTGCTGTCAGACCTGTATTAAGCTTTGGACGACA 417
|||||
134  sngLysnIleAlaGluGluLeuSerIleLeuLysTyrAsnThrAspSer 150
|||||
418  ATGGGAACATTGCTGAAGAACTGACCATTCANATGGAAACACAGACAGT 467
|||||
151  ValGluGluPheLeuSerGluLysLeuGluArgIle 162
|||||
468  GTAGAGAATTCCTGAGTGAAGTGGACGACATA 503
|||||

seq_name: gb_..:AF288740

seq_documentation_block:
LOCUS      AF288740              1463 bp      mRNA      linear      ROD 13-NOV-2000
DEFINITION Mus musculus 15 kda selenoprotein (Sep15) mRNA, complete cds.
ACCESSION  AF288740
VERSION    AF288740.1  GI:11139619
KEYWORDS
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 1463)
            Kumaraswamy, E., Malykh, A., Korotkov, K.V., Kozayvkin, S., Hu, Y.,
            Kwon, S.Y., Moustafa, M.E., Carlson, B.A., Berry, M.J., Lee, B.J.,
            Hatfield, D.L., Diamond, A.M. and Gladyshev, V.N.
            Structure-Expression Relationships of the 15-kda Selenoprotein

```

```

JOURNAL    J. Biol. Chem. 275 (45), 35540-35547 (2000)
REFERENCE  2 (bases 1 to 1463)
AUTHORS    Kumaraswamy, E., Malykh, A., Korotkov, K.V., Kozayvkin, S., Hu, Y.,
            Kwon, S.Y., Moustafa, M.E., Carlson, B.A., Berry, M.J., Lee, B.J.,
            Hatfield, D.L., Diamond, A.M. and Gladyshev, V.N.
TITLE      Direct Submission
JOURNAL    Submitted (21-JUL-2000) Department of Biochemistry, University of
            Nebraska, N51 Beadle Center, Lincoln, NE 68588, USA
FEATURES   Location/Qualifiers
            source          1..1463
                        /organism="Mus musculus"
                        /db_xref="taxon:10090"
            gene            1..1463
                        /gene="Sep15"
            CDS             247..735
                        /gene="Sep15"
                        /note="selenocysteine"
                        /codon_start=1
                        /transl_except="(pos:523..525,aa:OTHER)
                        /product="15 kda selenoprotein"
                        /protein_id="MAG31765.1"
                        /db_xref="GI:11139620"
            /translation="MAGPSGCVLPARGILRLILATYQANVARGAERSRACRELGFS
            SNLCSGCDLGGFNLPIDPDRGCCQEPQAEFTKLYAGAILVEGCGXKIGRPPOV
            AFVRSKPKLFRGLQIKYVGSDPVLKLDNDGNINAEILKMTNDSVEEFLSEKLE
            RI"
            misc_feature     1276..1371
                        /gene="Sep15"
                        /note="SECIS element: causes positions 523..525 (TGA) to
                        be translated as selenocysteine"

BASE COUNT      419 a      299 c      341 g      404 t
ORIGIN

alignment_scores:
    Quality: 747.00      Length: 162
    Ratio: 4.819      Gaps: 0
    Percent Similarity: 95.679      Percent Identity: 91.358

alignment_block:
US-09-676-718A-1 x AF288740

Align seg 1/1 to: AF288740 from: 1 to: 1463

1  MetaAlaIaGlyProSerGlyCysLeuValProAlaPheGlyLysArgLe 17
|||||
247  ATGGCGGCGAGGCGACGGTGGCTGCGCGCCGCTGCGGCTGCGCTT 296
|||||
17  uLeuLeuAlaThrValLeuGlnAlaValSerAlaPheGlyAlaGluPheS 34
|||||
297  GCTGTGCGACTGCGTTCAAGCGGCGTCTGCTTGGGGGAGAGATTG 346
|||||
34  eSerGluAlaCysArgGluLeuGlyPheSerSerAsnLeuLeuCysSer 50
|||||
347  CGTCAAGGCATGACAGAGAGTGGGTTCTCCAGCAACTGCTGTCAGC 396
|||||
51  SerCysAspLeuLeuGlyGlnPheAsnLeuGlnLeuAspProAspCys 67
|||||
397  TCTTGTGATCTTCTTGACAGTTCATCTGCTCCACAGTGGATCCTGTTG 446
|||||
67  sArgGlyCysCysGlnGluGlnAlaGlnPheGluThrLysLysLeuTYR 84
|||||
447  CAGAGGCTGCTGTGAGAAAGACAAATTGAAACCAAAAGCTGATG 496
|||||
84  IaGlyAlaIleLeuGlnValCysGly**LysLeuGlyArgPheProGln 100
|||||
497  CAGGAGCATCCTTGAAGTTCGCGATGAAATTTGGGAGGTTCCCTCA 546
|||||
101  ValGlnAlaPheValArgSerAspLysProLysLeuPheArgGlyLeuG1 117
|||||
547  GTCCAAAGCTTTTGTAGAGATGATAAACCAAACTTCTCAGAGCTTACA 596
|||||

```

117 nilelystyValArglySerAspProValLeuLysLeuLeuAspAsp 134
 597 GATCAAGTAACTTCAGAGCTCAGACCTCTACTAAAGCTTTGGACACACA 646
 134 snGlyAsnIleAlaGluGluLeuSerIleLeuLysTrpAsnThrAspSer 150
 647 ACGGGAACATGCTGAGAACTAAGCATCTCAAGTGAACACAGACAGT 696
 151 ValGluGluPheLeuSerGluLysLeuGluArgIle 162
 697 GTGGAAGAGTTCCTGAGCGGAGAGTGTGAACGACATA 732
 seq_name: gb_ro:BC019792
 seq_documentation_block:
 LOCUS BC019792 1238 bp mRNA linear ROD 22-JAN-2002
 DEFINITION Mus musculus, similar to 15-kDa selenoprotein, clone MGC:30351
 IMAGE:5003585, mRNA, complete cds.
 ACCESSION BC019792.1 GI:18044683
 VERSION BC019792.1 GI:18044683
 KEYWORDS MGC.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCES
 1 (bases 1 to 1238)
 Strausberg, R.
 Direct Submission
 Submitted (19-DEC-2001) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: http://mgc.ncl.nih.gov
 Contact: MGC help desk
 Email: cgapbs@mail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)
 DNA Sequencing by: Baylor College of Medicine Human Genome
 Sequencing Center
 Center code: BCM-HGSC
 Web site: http://www.hgsc.bcm.tmc.edu/cdna/
 Contact: amgebcm.tmc.edu
 Gunaratne, P.H., Garcia, A.M., Lu, X., Huylk, S.W., Hale, S.M.,
 Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
 Richards, S., Gibbs, R.A.
 Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LINL at: http://image.lnl.gov
 Series: IRAX Plate: 41 Row: J Column: 6
 This clone was selected for full length sequencing because it
 passed the following selection criteria: Similarity but not
 identity to protein.
 FEATURES
 source location/Qualifiers
 1..1238
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="MGC:30351 IMAGE:5003585"
 /tissue_type="Mammary tumor, MMTV-LTR/INT3 model, 5 month
 old mouse. Taken by biopsy."
 /clone_id="NCI_CGAP_Mam2"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"
 15..293
 /codon_start=1
 /product="Similar to 15-kDa selenoprotein"
 /protein_id="AAH19792.1"
 /db_xref="GI:18044684"
 /translation="MAAGGGWLRPALGIRLLIATFAQAVSALGAEPFSAECREIGFS
 SNLCSDDLGFQFNLPLDPVCRCCGEAEAFETKITYAGAILVECG"
 BASE COUNT 369 a 243 c 279 g 347 t
 ORIGIN

alignment_scores:
 Quality: 745.00 Length: 162
 Ratio: 4.806 Gaps: 0
 Percent Similarity: 95.679 Percent Identity: 91.358
 alignment_block:
 US-09-676-718a-1 x BC019792 ..
 Align seq 1/1 to: BC019792 from: 1 to: 1238
 1 MetAlaIaGlyProSerGlyCysLeuValProAlaPheGlyLysArgLe 17
 15 ATGGCGGCGAGGCGAGGCTGGCTGGCCCTCTGGGGCTGCCCTT 64
 17 uLeuLeuAlaThrValLeuGlnAlaValSerAlaPheGlyLysLeuPhe 34
 65 GCTGCTGGCGACCTCGCTTCAAGCGGTGTCTGCTCTGGGGGACAGTTTG 114
 34 eISerGluAlaCysArgGluLeuGlyPheSerSerAsnLeuLeuCysSer 50
 115 CGTCAGAGCGATGCAGAGAGTGGGTTCTCCACCACTGCTCTCGCAGC 164
 51 SerCysAspLeuLeuGlyGlnPheAsnLeuLeuGlnLeuAspProAsp 67
 165 TCTTGGCATCTCTTGGACAGATTAACTGCTCCACTGGACCTGTTTG 214
 67 sArgGlyCysGlyGlnGluGluAlaGlnPheGluThrLysLysLeuTyra 84
 215 CAGAGGCTGCTGTCAAGACAGACCAATTGGAACCAAAAGCTGTATG 264
 84 IaGlyAlaIleLeuGluValCysGly**LysLeuGlyArgPheProGln 100
 265 CAGAGCCATCTTGAAGTCTGCGATGAAATTTGGGAGGTGCTCTCA 314
 101 ValGlnAlaPheValArgSerAspLysProLysLeuPheArgGlyLeu 117
 315 GTCCAACTTTTGTCAAGAGATGAACCCAACTCTTCAAGAGTTCACA 364
 117 nilelystyValArglySerAspProValLeuLysLeuLeuAspAsp 134
 365 GATCAAGTAACTTCAGAGCTCAGACCTCTACTAAAGCTTTGGACACACA 414
 134 snGlyAsnIleAlaGluGluLeuSerIleLeuLysTrpAsnThrAspSer 150
 415 ACGGGAACATGCTGAGAACTAAGCATCTCAATGTAAGACACAGACAGT 464
 151 ValGluGluPheLeuSerGluLysLeuGluArgIle 162
 465 GTGGAAGAGTTCCTGAGCGGAGAGTGTGAACGACATA 500
 seq_name: gb_ro:AF390544
 seq_documentation_block:
 LOCUS AF390544 1236 bp mRNA linear ROD 16-JUL-2001
 DEFINITION Rattus norvegicus 15 kDa cytosolic selenoprotein mRNA, complete
 cds.
 ACCESSION AF390544
 VERSION AF390544.1 GI:14719271
 KEYWORDS
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 REFERENCES
 1 (bases 1 to 1236)
 Roethlein, D., Kyriakopoulos, A. and Behne, D.
 A 15 kDa-selenoprotein in several tissues of the rat
 Unpublished
 JOURNAL
 REFERENCES
 2 (bases 1 to 1236)
 Roethlein, D., Kyriakopoulos, A. and Behne, D.
 Direct Submission
 TITLE

JOURNAL Submitted (11-JUN-2001) SF 6, Hahn-Meitner-Institut, Glienicker
 Str. 100, Berlin D-14109, Germany

FEATURES
 source Location/Qualifiers
 1..1236
 /organism="Rattus norvegicus"
 /db_xref="taxon:10116"
 1..489
 /note="selenocysteine"
 /codon_start=1
 /transl_except="(pos:277..279,aa:OTHER)
 /product="15 kDa cytosolic selenoprotein"
 /protein_id="AAK73100.1"
 /db_xref="GI:14719272"
 /translation="MAAGCGGWLRLPGLRLATATQAVSALCAERSSRACRELGTS
 SNLSSCDLLGGFNLDPDPCVCGCCQEEAQQETKRLVAGALLVCGXKLRFPQVQ
 AFVRSDBPKLFRGLQIKYRGSVPVLLKLDNGNIABELSLKXNDTSVEFLSEKLE
 R"

misc_feature 1085..1132
 /note="selenocysteine insertion sequence (SECIS)"

polya_signal 1211..1216
 263 c 287 g 347 t 3 others

BASE COUNT 336 a 263 c 287 g 347 t 3 others
 ORIGIN

alignment_scores:
 Quality: 741.00 Length: 162
 Ratio: 4.812 Gaps: 0
 Percent Similarity: 95.062 Percent Identity: 91.975

alignment_block:
 US-09-676-718a-1 x AF390544 ..

Align seg 1/1 to: AF390544 from: 1 to: 1236

```

1 MetalaalaglProserglyCysleuValProalaphegliLysArgpe 17
|||||
1 AMGGGGGCGAGGCGGCGGTGCTCCGCCGCCGCTGCGCTTCTG 50
17 uleuLeuAlaThrValleuGlnAlaValSerAlaPheglYalagluPheS 34
|||||
51 ACTGTGCGACTGCGCTTCAAGCGGTGCTCTGCGGCGAGAGTCT 100
34 eSerGlnAlaCysArggluLeuGlyPheSerSerAsnleuLeuGlySer 50
|||||
101 COTCAGAGCATGCCGAGGAGTGGCTTCTCCAGCAACTGCTGTCAC 150
51 SerCysAspLeuLeuGlyGlnPheAsnleuLeuGlnLeuAspProAspCy 67
|||||
151 TCCGTGCGATCTCTTGAGACGTTTAACGTCTTCCACTGATCTGCTG 200
67 sArgGlyCysCysGlnGluAlaGlnPhegluThrLysLysLeuTyrA 84
|||||
201 CAGAGGCTGCTGTCAAGAGGAGCGCAGTTTGAACCAAAAAAGCTGTATG 250
84 laglValAlileuLeuGlnuValCysGly**LysLeuGlnYlAargPheProGln 100
|||||
251 CAGGAGCCATCTTGAAGTCTGTGATGAAATFTGGGGGTTCCCTCA 300
101 ValGlnAlaPheValArgSerAspLysProLysLeuPheArgGlyLeuGln 117
|||||
301 GTCCAAAGCTTTTGTCAAGAGCATTAACCAAAAGCTTTCAGAGGCTTACA 350
117 nileLysTyrValArgGlySerAspProValleuLysLeuLeuAspAspA 134
|||||
351 GATCAAGATATGTTCAAGGCTCAGACCTGTACTAAAGCTTTTGGAGACA 400
134 snGlyAsnIleAlaGluGluLeuSerIleLeuLysTyrAsnThrAspSer 150
|||||
401 ACGGAAATTTGCTGAAGAGCTCAGACATCTCAAGTDDAACACAGACAGT 450
151 ValGlnGluPheLeuSerGlyLysLeuGlnArgIle 162
|||||
451 GTGGAAGAGTTCTTGAAGAGAGAGTGAACGACAT 486

```

seq_name: gb:BC021697

seq_documentation_block:
 LOCUS BC021697 1249 bp mRNA linear PRI 18-JAN-2002
 DEFINITION Homo sapiens, clone MGC:24367 IMAGE:4051151, mRNA, complete cds.
 ACCESSION BC021697
 VERSION BC021697.1 GI:18203835
 KEYWORDS MGC.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 AUTHORS Strausberg, R.
 TITLE Direct Submission
 JOURNAL Submitted (14-JAN-2002) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
 NIH-MGC Project help desk
 Contact: MGC help desk
 Email: cgaps-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: CLONTECH Laboratories, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>
 contact: amadansystemsbiology.org
 Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha
 Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
 Series: IRAL Plate: 31 Row: 0 Column: 24
 This clone was selected for full length sequencing because it
 passed the following selection criteria: Hexamer frequency ORF
 analysis.

location/Qualifiers
 1..1249
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="MGC:24367 IMAGE:4051151"
 /tissue_type="kidney, hypernephroma"
 /clone_lib="NIH_MGC_58"
 /lab_host="DH10B"
 /note="Vector: pDNR-LIB"
 37..315
 /codon_start=1
 /product="Unknown (protein for MGC:24367)"
 /protein_id="AAH21697.1"
 /db_xref="GI:18203836"
 /translation="MAAGPSGLVPARGLRLLATVLAQAVSARFASRSRACRELGTS
 SNLSSCDLLGGFNLDPDPCVCGCCQEEAQQETKRLVAGALLVCG"

BASE COUNT 400 a 212 c 258 g 379 t
 ORIGIN

alignment_scores:
 Quality: 701.50 Length: 163
 Ratio: 4.906 Gaps: 2
 Percent Similarity: 87.730 Percent Identity: 87.117

alignment_block:
 US-09-676-718a-1 x BC021697 ..

Align seg 1/1 to: BC021697 from: 1 to: 1249

```

1 MetalaalaglProserglyCysleuValProalaphegliLysArgpe 17
|||||
37 AMGGGGGCGAGGCGGCGGTGCTGCGGCGGCTTGGGCTACGCT 86
17 uleuLeuAlaThrValleuGlnAlaValSerAlaPheglYalagluPheS 34

```



```

|||||
87 GTGTGGGCGACTGCTTCAGCGGTCTGCTTTGGGGGAGAGTTT 136
34 erserGluaIaCysArgGluLeuGlyPheSerSerAsnLeuLeuCysSer 50
|||||
137 CATCGGAGGATGACAGAGACTTAGGCTTTCTAGCAACTGCTTTGGACG 186
51 SerCysAspLeuLeuGlyGlnPheAsnLeuLeuGlnLeuAspProAspCy 67
|||||
187 TCTTGGATCTCTCGACAGTTCACCTGCTCAGCTGATGATCCTGATG 236
67 sArgGlyCysCysGlnGlnGlnGlnPheGlnThrLysLysLeuTyrA 84
|||||
237 CAGAGATGCTGTGAGAGAGACACCAATTTGAACCAAAAGCTGATG 286
84 laGlYAlaIleLeuGluValCysGly**LysLeuGlyArgPheProGln 100
|||||
287 CAGGACTATCTTGAAATTTGTGGATGAATAATTGGAGAGTTCCCT... 333
101 ValGlnAlaPheValArgSerAspLysProLysLeuPheArgGlyLeuG 117
334 .....CA 335
117 ntle.LysTyrValArgGlySerAspProValLeuLysLeuLeuAspAsp 133
|||||
336 AGTCAGATGATGTCGCTGTCAGACCTGTATTAAAGCTTTGGAGAC 385
134 AsnGlyAsnIleAlaGluGluLeuSerIleLeuLysTyrPasnThrAspSe 150
|||||
386 AATGGAGACATTCGTGAAGACTGACATCTCAATGGAACACAGACAG 435
150 rValGluGluPheLeuSerGluLysLeuGluArgIle 162
|||||
436 TGTAGAGAAATTCCTGAGTGAATAAGTTGAACCATTA 472
seq_name: gb_ro:BC010481

seq_documentation_block:
LOCUS BC010481 1020 bp mRNA linear ROD 12-JUL-2001
DEFINITION Mus musculus, clone IMAGE:3482046, mRNA, partial cds.
ACCESSION BC010481
VERSION BC010481.1 GI:14714679
KEYWORDS
SOURCE
ORGANISM
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS
1 (bases 1 to 1020)
Strausberg, R.
Direct Submission
Submitted (10-JUL-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-rt@mail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: villalob@bcm.tmc.edu.
Villalob, D.R., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,
A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,
Muzny, D.M., Gibbs, R.A.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAX Plate: 6 Row: D Column: 4
This clone was selected for full length sequencing because it

```

```

FEATURES
source
Location/Qualifiers
1..1020
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:3482046"
/tissue_type="Mammary tumor. MAP-TGF alpha model. 7 months
old, gross tissue."
/clone_id="NCI CGAP_Mam5"
/lab_host="DH10B"
/note="vector: pCMV-SPORT6"
<1..181
/codon_start=2
/product="Unknown (protein for IMAGE:3482046)."
/protein_id="AAH10481.1"
/db_xref="GI:14714680"
/translation="GAVYKKHKLKRSCKQEPKSLKADENMGSLKSLSEVINPNS
SEVYSSMEVQTLT"

BASE COUNT 340 a 188 c 199 g 293 t
ORIGIN

alignment_scores:
Quality: 476.00 Length: 94
Ratio: 5.118 Gaps: 0
Percent Similarity: 98.936 Percent Identity: 98.936

alignment_block:
US-09-676-718a-1 x BC010481 ..

Align seq 1/1 to: BC010481 from: 1 to: 1020
69 GTCyCysGlnGluAlaGlnPheGluThrLysLysLeuTyrAlaG 85
1 GGGTGGCTGTCAAGAGAAACCAATTTGAAACCAAAAGCTGTATGACAG 50
85 yAlaIleLeuGluValCysGly**LysLeuGlyArgPheProGlnValG 102
51 AGCCATCTCTGAAGCTGCGGATGAATAATTGGAGGTTCCCTCAAGTCC 100
102 lnaIaPheValArgSerAspLysProLysLeuPheArgGlyLeuGlnIle 118
101 AAGCTTTGTCAAGAGATGAATAAACCAACTCTTCAAGGCTTCAACATC 150
119 LysTyrValArgGlySerAspProValLeuLysLeuLeuAspAspAsnG 135
151 AAGTATGTTGAGGCTCAGACCCCTGTACTAAAGCTTTGGACGACACGG 200
135 yAsnIleAlaGluGluLeuSerIleLeuLysTyrPasnThrAspSerValG 152
201 GAACATTTGCGAAGAACTAAGCATCCCAATGGAACACAGACAGTGTGG 250
152 lUGluPheLeuSerGluLysLeuGluArgIle 162
251 AAGATTCCTGAGAGAGAGATTGGAACGCAATA 282
seq_name: gb_pat:AX330345

seq_documentation_block:
LOCUS AX330345 313 bp DNA linear PAT 09-JAN-2002
DEFINITION Sequence 854 from Patent WO0194629.
ACCESSION AX330345
VERSION AX330345.1 GI:18103323
KEYWORDS
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.
REFERENCE
AUTHORS
1 (sites)
Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,
Horrigan, S., Soppet, D.R. and Weaver, Z.
Cancer gene determination and therapeutic screening using signature

```

gene sets
JOURNAL Patent: WO 0194629-A 854 13-DEC-2001;
FEATURES Avalon Pharmaceuticals (US)
Location/Qualifiers
Source 1. .313
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 64 a 61 c 92 g 96 t
ORIGIN

alignment_scores:
Quality: 392.00 Length: 103
Ratio: 4.455 Gaps: 2
Percent Similarity: 85.437 Percent Identity: 84.466

alignment_block:
US-09-676-718A-1 x AX330345 ..

Align seg 1/1 to: AX330345 from: 1 to: 313

9 LeuValProAlaPheGlyLysArgLeuLeuAlaThrValLeuGlnAl 25
|||||
1 CTGGTGGCCGGCTTGGGCTAGCTGTTGGCTGGGACCTGCTTCAAGC 50
25 aValSerAlaPheGlyAlaGlnPheSerSerGlnAlaCysArgGlnLeuG 42
|||||
51 GTTGTCTGCTTTGGGCGAGATTTTCATCGAGGACATCGAGAGATTAG 100
42 LYPheSerSerAsnLeuLeuCysSerSerCysAspLeuLeuGlnPhe 58
|||||
101 GCTTTTTCAGACACTGCTTTGACACTCTTGATCTTCTCGGACAGTTC 150
59 AsnLeuLeuGlnLeuAspProAspCysArgGlyCysGlnGlnGlnAl 75
|||||
151 AACCTGCTTACGACCTGATCTGATTCAGAGATGCTGTCAGAGAGAGC 200
75 aGlnPheGlnThrLysLysLeuTyraGlyAlaAlaLeuLeuGlnAlaCysG 92
|||||
201 ACAATTGTAACCAAA.AAGCTGTAAGCAGAGCTATTCTTGAGTTTGTG. 248
92 Lys**LysLeuGlyArgPhe.....Pro 99
|||||
249 ..GAPTAATTGGGAAGTTCCTCAGTCACGCTTGTAGGGTGATTAACA 296
100 GlnValGln 102
|||||
297 ACTGTTTCAG 305

seq_name: gb_in:AF457547

seq_documentation_block:
LOCUS AF457547 624 bp mRNA linear INV 29-JAN-2002
DEFINITION Anopheles gambiae selenoprotein mRNA, partial cds.
ACCESSION AF457547
VERSION AF457547.1 GI:18389880
KEYWORDS
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
Culicoidae; Anophelinae;
1 (bases 1 to 624)
TOWARDS a catalog for genes and proteins from the salivary gland of
the malaria vector, Anopheles gambiae
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 624)
AUTHORS Francischetti,I.M., Valenzuela,J.G. and Ribeiro,J.M.
TITLES Towards a catalog for genes and proteins from the salivary gland of
the malaria vector, Anopheles gambiae
JOURNAL Francischetti,I.M., Valenzuela,J.G. and Ribeiro,J.M.
AUTHORS Direct Submission
TITLES Submitted (12-DEC-2001) Section of Medical Entomology/Laboratory of
Parasitic Diseases/NIH/ID, National Institutes of Health, 4 Center
Drive, Building 4, Room 126, MSC 0425, Bethesda, MD 20892-0425, USA

FEATURES Location/Qualifiers
Source 1. .624
/organism="Anopheles gambiae"
/db_xref="taxon:7165"
CDS <1..492
/note="salivary protein"
/codon_start=1
/product="selenoprotein"
/protein_id="FAL68777.1"
/db_xref="GI:18389881"
SDYGLIEKHEKCECKDKDTKADSKLVYPAVALVECTKFGAVPQIQATIKSDRPK
FPMITIKYVRGLDPIYALMDQGVETLISINMNTDFTVOEFETRLAKVEDDYIKT
NRV"
BASE COUNT 213 a 118 c 148 g 145 t
ORIGIN

alignment_scores:
Quality: 383.00 Length: 151
Ratio: 3.165 Gaps: 3
Percent Similarity: 80.132 Percent Identity: 48.344

alignment_block:
US-09-676-718A-1 x AF457547 ..

Align seg 1/1 to: AF457547 from: 1 to: 624

16 ArgLeuLeuAlaThrValLeu.....GlnAlaValSerAlaPheG1 30
|||||
4 CACACTTTGCCATTACGTCGTCTTATTTCGATTTGATGACGTGATTGG 53
30 YAlaGlnPheSerSerGlnAlaCysArgGlnLeuGly...PheSerSera 46
|||||
54 TGCACAAATTCTCCGCCGAAGATTGCCGGAGCTGGGGCTTATCAATGCC 103
46 snLeuLeuCysSerSerCysAspLeuLeuGlnGlnPheAsnLeuLeuGln 62
|||||
104 AACCTGTTTGTTCGGCGGTGTAGTACTTAAGCATTAAGATGATGATCA 153
63 LeuAspProAspCysArgGlyCysCysGlnGlnGlnAlaGlnPheGln 79
|||||
154 CTGAAGAAGACATTGCTTGAATGTTGCCAAAGACACTGAAGCCGATTC 203
79 Lys...LysLeuTyraGlyAlaAlaLeuLeuGlnAlaCysGly**LysL 95
|||||
204 AAAACTGAAGGTGATCCGGCGGAGTGTGAGAGGTGTCAGAGTCCAGT 253
95 euGlyArgPheProGlnValGlnAlaPheValArgSerAspLysProLys 111
|||||
254 TTGGCGCATATCCACAGATTCAGGGCTTCATTAAAGGACCGACCGGCC 303
112 LeuPheArgGlyLeuGlnLleLysTyraValArgGlySerAspProVal 128
|||||
304 AAGTTTCCCAACCTAACCATCAAGTACGTGCTGTTGGATCCATCGT 353
128 uLysLeuLeuAspAspAsnGlyAsnLleAlaGlnLleuSerLleuLys 145
|||||
354 GAAGCTAATGACGACGAGGACGACGTGAAGGAACGCTTCCATTAACA 403
145 yStrPAsnThrAspSerValGlnGlnPheLeuSerGlnLysLeuGlnArg 161
|||||
404 AATGGACACCGACACGGGTGAGAGATTTCGAAACACGTTTGGCGAAA 453
162 Ile 162
|||||
454 GTT 456

seq_name: gb_pr:AF288992

seq_documentation_block:
LOCUS AF288992 64165 bp DNA linear PRI 12-NOV-2000
DEFINITION Homo sapiens 15 kDa selenoprotein (SEPI5) gene, complete cds.

ACCESSION AF288992
 VERSION AF288992.1 GI:11138956
 KEYWORDS
 SOURCE
 ORGANISM human.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 64165)
 Kumaraswamy, E., Maljkh, A., Korotkov, K.V., Kozayvkin, S., Hu, Y.,
 Kwon, S.Y., Moustafa, M.E., Carlson, B.A., Berry, M.J., Lee, B.J.,
 Hatfield, D.L., Diamond, A.M. and Gladyshev, V.N.
 Hatfield, D.L., Diamond, A.M. and Gladyshev, V.N.
 Structure-Expression Relationships of the 15-kDa Selenoprotein
 Gene. POSSIBLE ROLE OF THE PROTEIN IN CANCER ETIOLOGY
 J. Biol. Chem. 275 (45), 35540-35547 (2000)
 TITLE 2 (bases 1 to 64165)
 Kumaraswamy, E., Maljkh, A., Korotkov, K.V., Kozayvkin, S., Hu, Y.,
 Kwon, S., Moustafa, M.E., Carlson, B.A., Berry, M.J., Lee, B.J.,
 Hatfield, D.L., Diamond, A.M. and Gladyshev, V.N.
 Direct Submission
 Submitted (24-JUL-2000) Department of Biochemistry, University of
 Nebraska, N151 Beadle Center, Lincoln, NE 68588, USA
 FEATURES
 Location/Qualifiers
 1..64165
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 join(6777..6873,17484..17651,40213..40276,52814..52863,
 57310..58192)
 /gene="SEPI5"
 /product="15 kDa selenoprotein"
 6777..6798
 /gene="SEPI5"
 6777..58192
 /gene="SEPI5"
 6777..6873
 /gene="SEPI5"
 /number=1
 join(6799..6873,17484..17651,40213..40276,52814..52863,
 57310..57441)
 /gene="SEPI5"
 /note="contains SECIS element; selenocysteine"
 /codon_start=1
 /transl_except=(pos:40246..40248,aa:OTHER)
 /product="15 kDa selenoprotein"
 /protein_id="AA031557.1"
 /db_xref="GI:11138957"
 /translation="MAAGPSGCLVPAPFGRLRLATVIVQAVSAFGAFESSEACRELIGFS
 SNLACSSCDLIGQFNLLQLDPCRCCEAEAFETKKLYAGAILLEVCGXKIGRFPQV
 AFVRSDDRKLFRGLQIKYVRGSDPYLKLDDNGNIAELSLIKWNTDVEEFLSEKLE
 RI"
 17484..17651
 /gene="SEPI5"
 /number=2
 17652..40212
 /gene="SEPI5"
 /note="contains differences with sequence deposited in
 Genbank Accession Number AL121989"
 /number=2
 40213..40276
 /gene="SEPI5"
 /number=3
 52814..52863
 /gene="SEPI5"
 /number=4
 57310..58192
 /gene="SEPI5"
 /number=5
 57442..58192
 /gene="SEPI5"
 3' UTR 57442..58192
 BASE COUNT 18805 a 12503 c 12005 g 20852 t
 ORIGIN

alignment_scores:
 Quality: 304.00 Length: 135
 Ratio: 3.102 Gaps: 6
 Percent Similarity: 72.593 Percent Identity: 54.815
 alignment_block:
 US-09-676-718a-1 x AF288992 ..
 Align seg 1/1 to: AF288992 from: 1 to: 64165
 26 valSerAlaPheGlyAlaGluPheSerSerGluAlaCysArgGluLeuG1 42
 17484 GTGCTGCTTGTGGGACAGAGTTTTCATCGACGACATGACAGAGTTAGG 17533
 42 yPheSerSerAsnLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 59
 17534 CTTTCTTACGACACTGCTTGGACGCTCTTGATCTTCTCGACACTTCA 17583
 59 snLeuLeuGlnLeuAspProAspCysArgGlyCysGlnGluAla 75
 17584 ACCTGCTTCAGCTGATCCTGATTCAGAGAGATGCTGTGAGAGACA 17633
 76 GlnPheGlnThrLysLysLeuTyraGlyAlaIleLeuGlnValCysG1 92
 17634 CAATTGAAACCAAAAAGTACTGTTTCTTATGATA..... 17669
 92 y***LysLeuGlyArgPheProGlnValGlnAlaPheValArgSerAsp. 108
 17670TTCCCTCCATTTGATGTTTATGATGACCAAGGT 17703
 109 ..LysProLysLeuPheArgGlyLeuGlnIleLysTyraValArgGlySer 124
 17704 AAAAATGTAAGTTCCTTAAT.....CAATTAAAGTAAAAAGCTCT 17747
 125 AspProValLeuLysLeuLeuAspAspAsnGlyAsnIleAlaGluLeu 141
 17748AAGCTTTTA.....CATTTTATGATAAAACT 17773
 141 user...IleLeuLysTrpAsnThrAspSerValGluGluPheLeuSer 157
 17774 CAACGTATTTCTTGTCTGCGCTATGATTCATCAACAAATATGTAAG 17823
 157 LuLys 158
 17824 AGAAG 17828
 seq_name: gp_hcg:AL355981
 seq_documentation_block:
 LOCUS AL355981 155871 bp DNA linear HTG 23-JAN-2001
 DEFINITION Homo sapiens chromosome 1 clone RP11-277G24, *** SEQUENCING IN
 PROGRESS ***, 15 unordered pieces.
 ACCESSION AL355981 GI:9797397
 VERSION AL355981.3
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 155871)
 Plumb, B.
 TITLE Direct Submission
 JOURNAL Submitted (21-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 requests: clonerequest@sanger.ac.uk
 On Aug 12, 2000 this sequence version replaced gi:92133687.
 COMMENT
 ----- Genome Center
 Center: Sanger Centre
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: humquery@sanger.ac.uk

```

----- Project Information
Center project name: ha277624
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 146127 bases at least Q40
Consensus quality: 149488 bases at least Q30
Consensus quality: 151696 bases at least Q20
Insert size: 154471; sum-of-contigs
Insert size: 176463; 14.5% error; agarose-fp
Quality coverage: 3.58x in Q20 bases; sum-of-contigs Quality
coverage: 3.19x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 15247: contig of 15247 bp in length
15248 15347: gap of 100 bp
15348 21318: contig of 5971 bp in length
21319 21418: gap of 100 bp
21419 25299: contig of 3881 bp in length
25300 25399: gap of 100 bp
25400 30376: contig of 4977 bp in length
30377 30476: gap of 100 bp
30477 33263: contig of 2787 bp in length
33264 33363: gap of 100 bp
33364 72796: contig of 39433 bp in length
72797 72896: gap of 100 bp
72897 98426: contig of 25530 bp in length
98427 98526: gap of 100 bp
98527 106876: contig of 8350 bp in length
106877 106976: gap of 100 bp
106977 111928: contig of 4952 bp in length
111929 112028: gap of 100 bp
112029 119408: contig of 7380 bp in length
119409 119508: gap of 100 bp
119509 137994: contig of 18486 bp in length
137995 138094: gap of 100 bp
138095 142132: contig of 4038 bp in length
142133 142332: gap of 100 bp
142333 144266: contig of 2034 bp in length
144267 144366: gap of 100 bp
144367 148940: contig of 4574 bp in length
148941 149040: gap of 100 bp
149041 155871: contig of 6831 bp in length.
Location/Qualifiers
1. 155871
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone_1lb="RPCT-11.1"
1. 15247
/note="assembly_fragment:00782
fragment_chain:1"
15348. 21318
/note="assembly_fragment:01104
fragment_chain:1"
21419. 25299
/note="assembly_fragment:00085"
25400. 30376
/note="assembly_fragment:00125"
30477. 33263
/note="assembly_fragment:00175"
33364. 72796
/note="assembly_fragment:00232"
72897. 98426

```

```

misc_feature /note="assembly_fragment:00364"
98527. 106876
misc_feature /note="assembly_fragment:00437"
106977. 111928
misc_feature /note="assembly_fragment:00459"
112029. 119408
misc_feature /note="assembly_fragment:00503"
119509. 137994
misc_feature /note="assembly_fragment:00572"
138095. 142132
misc_feature /note="assembly_fragment:00752"
142233. 144266
misc_feature /note="assembly_fragment:01162"
144367. 148940
misc_feature /note="assembly_fragment:01243"
149041. 155871
misc_feature /note="assembly_fragment:01367"
BASE COUNT 44562 a 31293 c 3116 g 47497 t 1403 others
ORIGIN
alignment_scores:
Quality: 304.00 Length: 135
Ratio: 3.102 Gaps: 6
Percent Similarity: 72.593 Percent Identity: 54.815
alignment_block:
US-09-676-718A-1 x AL355981 ..
Align seg 1/1 to: AL355981 from: 1 to: 155871
26 valserAlaPheGlyAlaGluPheSerSerGluAlaCysArgGluLeuG1 42
38770 GGTCTGCTTTGGGGCAGAGTTTCATCGAGCGATCAGAGATTGAG 38819
42 yPheSerSerAlaLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 59
38870 CTTTCTTGACCACTGCTTTCGAGCTCTTGATCTTTCGACACGTTCA 38869
59 snLeuLeuGlnLeuAspProAspCysArgGlyCysCysGlnGluAla 75
38870 ACCTCCTTCAGCTGATGCTGATTCAGAGATGCTGTCAGAGAGAACCA 38919
76 GlnPheGluThrLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 92
38920 CAATTGAAACCAAAAGCTACTGTTTCTAGTATA..... 38955
92 y**LysLeuGlyArgPheProGlnValGlnAlaPheValArgSerAsp. 108
38956 .....TTCCTTCCTTGATGATGCTTATGATGACACAGCT 38989
109 ..LysProLysLeuPheArgGlyLeuGlnLeuLeuLeuLeuLeuLeu 124
38990 AAAATGTAAGTGTGCTTAAT.....CAATTAACTGTAACCAACCTCT 39033
125 AsProValLeuLysLeuLeuAspAspAsnGlnLeuLeuLeuLeuLeu 141
39034 .....AACCTTTTA.....CAATTGTAAGTAAACCT 39059
141 user...IleLeuLysTrpAsnThrAspSerValGluGluPheLeuSerg 157
39060 CACACGATTCTCTGCTGCGCATTAATGATCAATCAACCAATATGAAGG 39109
157 LuLys 158
39110 AGAAG 39114
seq_name: gb_Dr:AF267982S2
seq_documentation_block:
LOCUS AF267982S2 1743 bp DNA linear PRI 27-JUN-2000
DEFINITION Homo sapiens 15 kda selenoprotein gene, exon 2.
ACCESSION AF267983

```

```

VERSION      AF267983.1  GI:87455559
KEYWORDS
SEGMENT      2 of 5
SOURCE        human.
ORGANISM      Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 1743)
AUTHORS       Ryu, M. and Moon, E.
TITLE         The human 15-kDa selenoprotein gene : Characterisation of the
              genomic structure and functional analysis of the promoter
JOURNAL       Unpublished
REFERENCE     2 (bases 1 to 1743)
AUTHORS       Ryu, M. and Moon, E.
TITLE         Direct Submission
JOURNAL       Submitted (16-MAY-2000) Biological Science, Ajou University, San-5,
              Wonchun-dong, Paldal-gu, Suwon, Kyungki-do 442-749, South Korea
FEATURES
  source       1..1743
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone="2075D19"
  exon         795..962
               /number=2
  BASE COUNT   534 a      312 c      342 g      540 t      15 others
  ORIGIN
26 valSerAlaPheGluAlaGluPheSerSerGluAlaCysArgGluLeu 42
42 yPheSerSerAsnLeuLeuCysSerSerCysAspLeuLeuGlyGlnPhe 59
845 CTTTCTTACCACTTCTTGGCAGCTTGTGATCTTCTCGGACAGTTCA 894
59 snLeuLeuGlnLeuAspProAspCysArgGlyCysGlnGluAla 75
895 ACCTGCTTCAGCTGATCTGATTCAGAGAGATGCTGTCAGGAGAACA 944
76 GlnPheGluThrLysLys...LeuTyAlaGlyAlaIleLeuGluValCy 91
945 CATTGGAAACCAAAAGGACTGTTTCTAGATATTCCTTCATTGGA 994
91 scLy 92
995 TGGT 998
seq_name: gb_pr:HSDJ604K5
seq_documentation_block:
LOCUS      HSDJ604K5      127201 bp      DNA      linear      PRI 24-NOV-2000
DEFINITION Human DNA sequence from clone RP4-604K5 on chromosome 1p22.2-31.1.
              Contains a gene for a 15 kDa selenoprotein and a part of the
              heparan sulfate 2-sulfotransferase gene (K1AA0448). Contains ESTs,
              STSs, GSSs and Cpg Islands, complete sequence.
ACCESSION   ALI21989.12  GI:8919105
VERSION     ALI21989.12  GI:8919105
KEYWORDS    HTG: Cpg Island; K1AA0448; selenoprotein; sulfotransferase.
SOURCE      human.
ORGANISM    Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

```

REFERENCE     1 (bases 1 to 127201)
AUTHORS       Pearce, A.
TITLE         Direct Submission
JOURNAL       CB10 15A, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone
              requests: clonerequests@sanger.ac.uk
              On Jul 5, 2000 this sequence version replaced gi:8894634.
              During sequence assembly data is compared from overlapping clones.
              Where differences are found these are annotated as variations
              together with a note of the overlapping clone name. Note that the
              variation annotation may not be found in the sequence submission
              corresponding to the overlapping clone, as we submit sequences with
              only a small overlap as described above.
              This sequence has been finished according to sequence map criteria
              as follows. An attempt is made to resolve all sequencing problems,
              such as compressions and repeats, but not necessarily within known
              annotated human repeat sequence elements (e.g. Alu). Where the
              sequence is ambiguous, there is an annotation using the 'unsure'
              feature key.
              The following abbreviations are used to associate primary accession
              numbers given in the feature table with their source databases:
              Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
              on the WORMPEP database can be found at
              http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
              was generated from part of bacterial clone contigs of human
              Chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
              Group. Further information can be found at
              http://www.sanger.ac.uk/HGP/Chr1
              RP4-604K5 is from the library RPCI-4 constructed at the Roswell
              Park Cancer Institute by the group of Pieter de Jong. For further
              details see http://bacpac.med.buffalo.edu/
              VECTOR: pCYPAC2
              This sequence is the entire insert of clone RP4-604K5.
FEATURES
  source       1..127201
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /chromosome="1"
               /map="p22.2-31.1"
               /clone="RP4-604K5"
               /clone_id="RPCI-4"
  repeat_region 1..381
               /note="MUTID repeat: matches 113..504 of consensus"
               /note="74"
               /note="LM1M4 repeat: matches 7364..7597 of consensus"
               /note="843..925"
               /note="LM1M4 repeat: matches 7888..7977 of consensus"
               /note="874..1348"
               /note="match: GSS: Em:A0432443"
               /note="1621..1921"
               /note="Alu1 repeat: matches 1..299 of consensus"
               /note="1951..1982"
               /note="MIR repeat: matches 117..148 of consensus"
               /note="2021..2070"
               /note="Alu repeat: matches 253..302 of consensus"
               /note="2138..2181"
               /note="L2 repeat: matches 2648..2691 of consensus"
               /note="2282..2379"
               /note="L2 repeat: matches 2582..2693 of consensus"
               /note="2483..2769"
               /note="AluSc repeat: matches 1..287 of consensus"
               /note="2839..2962"
               /note="MIR repeat: matches 20..170 of consensus"
               /note="3136..3434"
               /note="Alu repeat: matches 4..302 of consensus"
               /note="complement(3237..3650)"
               /note="match: GSS: Em:AQ204224"
               /note="3657..4233"
               /note="match: GSS: Em:AQ376380"
               /note="3668..4152"
               /note="match: GSS: Em:AQ697506"
               /note="4591..4712"
               /note="L2 repeat: matches 2575..2710 of consensus"
  repeat_region repeat_region
  misc_feature   misc_feature
  misc_feature   misc_feature
  repeat_region  repeat_region

```

```
repeat_region 5034..5114 /note="L2 repeat: matches 425..516 of consensus"
repeat_region 5135..5447 /note="LRR3 repeat: matches 1..313 of consensus"
repeat_region 5636..5822 /note="AluSq repeat: matches 9..206 of consensus"
repeat_region 5926..6212 /note="LRR33 repeat: matches 1..297 of consensus"
repeat_region 6258..6562 /note="AluY repeat: matches 1..305 of consensus"
repeat_region 6670..6805 /note="AluSc repeat: matches 1..153 of consensus"
repeat_region 6896..7102 /note="MIR repeat: matches 32..262 of consensus"
repeat_region 7619..7813 /note="MER58A repeat: matches 36..224 of consensus"
repeat_region 8058..8357 /note="AluSq repeat: matches 1..300 of consensus"
repeat_region 9236..9365 /note="FLAM_C repeat: matches 1..129 of consensus"
repeat_region 9437..9569 /note="AluY repeat: matches 3..137 of consensus"
repeat_region 9570..9867 /note="AluY repeat: matches 1..297 of consensus"
repeat_region 9868..10039 /note="AluY repeat: matches 137..303 of consensus"
repeat_region 10040..10267 /note="L2 repeat: matches 2524..2747 of consensus"
repeat_region 10491..10777 /note="AluSp repeat: matches 1..288 of consensus"
repeat_region 10941..11506 /note="match: GSS: Em:AQ240745"
repeat_region 10964..11765 /note="match: GSS: Em:AQ385701"
repeat_region 11010..11339 /note="MIR4D repeat: matches 216..981 of consensus"
repeat_region 12369..12666 /note="match: GSS: Em:AQ549616"
repeat_region 13196..13505 /note="AluSq repeat: matches 1..300 of consensus"
repeat_region 13506..13644 /note="L2 repeat: matches 1260..1439 of consensus"
repeat_region 14377..14548 /note="match: GSS: Em:AQ233950"
repeat_region 14808..15098 /note="MIR repeat: matches 22..191 of consensus"
repeat_region 15313..15407 /note="MIRB repeat: matches 1..291 of consensus"
repeat_region 15408..15708 /note="AluSc repeat: matches 10..109 of consensus"
repeat_region 15709..16030 /note="AluSc repeat: matches 5..305 of consensus"
repeat_region 16038..16338 /note="MIRB repeat: matches 109..390 of consensus"
repeat_region 16360..16819 /note="L2 repeat: matches 1316..1657 of consensus"
repeat_region 16819..16819 /note="match: GSS: Em:AQ474120"
repeat_region 16803..16803 /note="match: GSS: Em:AQ469284"
repeat_region 17101..17411 /note="L2 repeat: matches 1674..1816 of consensus"
repeat_region 17557..17734 /note="AluX repeat: matches 1..312 of consensus"
```

```
repeat_region /note="L2 repeat: matches 1995..2197 of consensus"
repeat_region 17741..17772 /note="MER5A repeat: matches 1..34 of consensus"
repeat_region 17773..18080 /note="AluY repeat: matches 1..306 of consensus"
repeat_region 18081..18252 /note="MER5A repeat: matches 34..189 of consensus"
repeat_region 18922..19032 /note="AluSp/q repeat: matches 201..311 of consensus"
repeat_region 19609..19843 /note="MSTA repeat: matches 198..426 of consensus"
repeat_region 19844..20142 /note="AluY repeat: matches 1..297 of consensus"
repeat_region 20143..20338 /note="MSTB repeat: matches 1..198 of consensus"
repeat_region 20545..20913 /note="MSTB repeat: matches 1..198 of consensus"
repeat_region 21123..21418 /note="MSTB repeat: matches 17..391 of consensus"
repeat_region 21688..22106 /note="AluSq repeat: matches 1..297 of consensus"
repeat_region 22107..22383 /note="MIRG repeat: matches 10..423 of consensus"
repeat_region 22384..22405 /note="AluY repeat: matches 13..287 of consensus"
repeat_region 22406..22532 /note="MIRG repeat: matches 423..443 of consensus"
repeat_region 22533..22841 /note="MER30 repeat: matches 1..125 of consensus"
repeat_region 22842..22931 /note="AluY repeat: matches 1..306 of consensus"
repeat_region 22932..23242 /note="MER30 repeat: matches 125..223 of consensus"
repeat_region 23252..23300 /note="AluSq repeat: matches 1..311 of consensus"
repeat_region 23477..23777 /note="MIRG repeat: matches 443..506 of consensus"
repeat_region 24394..24690 /note="AluX repeat: matches 5..299 of consensus"
repeat_region 24702..24820 /note="AluSq repeat: matches 1..307 of consensus"
repeat_region 24821..25108 /note="L1M3 repeat: matches 5992..6105 of consensus"

alignment_scores:
  Quality: 303.50 Length: 68
  Ratio: 4.742 Gaps: 1
Percent Similarity: 94.118 Percent Identity: 86.765

alignment_block:
US-09-676-718a-1 x HSDJ604K5/rev ..
Align seg 1/1 to reverse of: HSDJ604K5 from: 1 to: 127201
26 ValSerAlaPheGluAlaGluPheSerSerGluAlaCysArgGluLeuL 42
70891 GTGTGCTTTGGGCAAGATTTCATCGAGGATCGAGGAGTATAGG 70842
42 yPheSerSerAlaLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 59
70841 CTTTTCATGCAACCTGCTTTCAGCTCTTGATGATCTTTCGAGACGTTCA 70792
59 snLeuLeuGlnLeuAspProAspCysArgGlyCysCysGlnGluAla 75
70791 ACCTTCCTTCAGCTGATCTGATTCAGAGGATGCTGTCAGAGGAGCA 70742
76 GlnPheGluThrLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuVal 91
70741 CAATTGAAACCAAAAGTACTGTTTCTAGATATATCTTCCTCCATTGCA 70692
91 sGly 92
70691 TGCT 70688
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2002, 16:19:15 ; Search time 66.54 Seconds

(without alignments)
421.178 Million cell updates/sec

Title: US-09-676-718a-1

Perfect score: 831

Sequence: 1 MAAGPSCGLVPAGFKRLLLA.....ILKMWDSVEEFLSEKLRI 162

Scoring table: BLOSUP62

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL_19:**
1: sp_archaea:**
2: sp_bacteria:**
3: sp_fungi:**
4: sp_human:**
5: sp_invertebrate:**
6: sp_mammal:**
7: sp_mhc:**
8: sp_organelle:**
9: sp_phage:**
10: sp_plant:**
11: sp_rodent:**
12: sp_virus:**
13: sp_vertebrate:**
14: sp_unclassified:**
15: sp_virus:**
16: sp_bacteriophage:**
17: sp_archaeal:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	822	98.9	162	4	Q9GZM0
2	750	90.3	162	11	Q9ERR7
3	744	89.5	162	11	Q923V8
4	474	57.0	92	4	Q9BS64
5	311	37.4	142	5	Q9VVJ7
6	250.5	30.1	152	5	Q9N4C6
7	226.5	27.3	158	10	Q9SVL3
8	81	9.7	68	10	Q22639
9	77	9.3	523	5	Q9VQ89
10	76.5	9.2	653	5	Q94594
11	76.5	9.2	653	5	Q25292
12	76.5	9.2	664	10	Q39830
13	76	9.1	265	16	P73692
14	75.5	9.1	220	5	Q94785
15	75.5	9.1	261	5	Q03929
16	75.5	9.1	316	10	Q9FHK0

17	75.5	9.1	653	5	Q26936	Q26936 trypanosoma
18	75.5	9.1	678	5	Q26937	Q26937 trypanosoma
19	75	9.0	456	12	Q993Y0	Q993Y0 spirella yel
20	74.5	9.0	665	10	Q94IK4	Q94IK4 schertfella
21	74.5	9.0	704	13	Q73774	Q73774 gallus gall
22	73	8.8	273	10	Q9M2U5	Q9M2U5 arabidopsis
23	73	8.8	274	10	Q23248	Q23248 arabidopsis
24	72.5	8.7	205	5	Q95Y00	Q95Y00 caenorhabdit
25	72.5	8.7	762	10	Q9LE55	Q9LE55 oryza sativ
26	72.5	8.7	830	5	Q93176	Q93176 caenorhabdit
27	72.5	8.7	838	5	Q9U919	Q9U919 caenorhabdit
28	72.5	8.7	838	5	Q95QY9	Q95QY9 caenorhabdit
29	72.5	8.7	1192	10	Q9LYN8	Q9LYN8 arabidopsis
30	72	8.7	383	3	Q01678	Q01678 pneumocystis
31	72	8.7	446	5	Q60996	Q60996 trypanosoma
32	72	8.7	620	5	Q9NGM4	Q9NGM4 drosophila
33	72	8.7	665	10	Q9M4E8	Q9M4E8 drosophila
34	72	8.7	777	10	Q9M0X6	Q9M0X6 cucumis sat
35	71.5	8.6	432	5	Q9N6X2	Q9N6X2 drosophila
36	71.5	8.6	514	5	Q94F41	Q94F41 leishmania
37	71.5	8.6	514	5	Q76958	Q76958 leishmania
38	71.5	8.6	524	10	Q91GX9	Q91GX9 oryza sativ
39	71.5	8.6	663	5	Q9GXR3	Q9GXR3 leishmania
40	71.5	8.6	829	5	Q9UB06	Q9UB06 drosophila
41	71.5	8.6	872	5	Q9NGQ0	Q9NGQ0 drosophila
42	71.5	8.6	1175	5	Q9YX8	Q9YX8 drosophila
43	71.5	8.6	1175	5	Q9VCX2	Q9VCX2 drosophila
44	71	8.5	421	10	Q92PR4	Q92PR4 arabidopsis
45	71	8.5	560	3	Q94667	Q94667 schizosacch

ALIGNMENTS

RESULT	ID	Q9GZM0	PRELIMINARY:	PRT:	162 AA.
Q9GZM0	AC	Q9GZM0	Q9GZM0		
DT	01-MAR-2001	(TREMBLrel. 16, Created)			
DT	01-MAR-2001	(TREMBLrel. 16, last sequence update)			
DT	01-MAR-2001	(TREMBLrel. 16, last annotation update)			
DE	DJ604K5.1	(15 KDA SELENOPROTEIN).			
GN	DJ604K5.1	OR SEP15.			
OS	Human sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]	SEQUENCE FROM N.A.			
RP	Pearce A.;				
RA	Submitted (Aug-2000) to the EMBL/Genbank/DBJ databases.				
RL	[2]				
RN	SEQUENCE FROM N.A.				
RP	Pubmed=10945981;				
RA	Kumaraswamy E., Malynk A., Korotkov R.V., Kozayavkin S., Hu Y.,				
RA	Kwon S.Y., Moustafa M.E., Carlson B.A., Berry M.J., Lee B.J.,				
RA	Hatfield D.L., Diamond A.M., Gladyshev V.N.;				
RT	POSSIBLE ROLE OF THE PROTEIN IN CANCER ETIOLOGY.;				
DR	EMBL; AL121989; CAC04186.1; -				
DR	EMBL; AF288991; AAG31556.1; -				
SO	SEQUENCE 162 AA; 17750 MW; 5894FBB423A089EE CRC64;				

Query Match 98.9%; Score 822; DB 4; Length 162;
Best Local Similarity 99.4%; Pred. No. 6.4e-79;
Matches 161; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAAGPSCGLVPAGFKRLLLATVLAQASAFGEPSSEACRELGSSNLLCSCDILGQFNL 60
DB 1 MAAGPSCGLVPAGFKRLLLATVLAQASAFGEPSSEACRELGSSNLLCSCDILGQFNL 60

QY 61 LQIDPDCGCCGCEAOFETKTLVAGAILLEVCGXKLGRRPVOAVRSDPKPLFRGLQIKY 120
 DB 61 LQIDPDCGCCGCEAOFETKTLVAGAILLEVCGXKLGRRPVOAVRSDPKPLFRGLQIKY 120
 QY 121 VRGSDPVKLKLDNGNIAEELSLIKWNTDSVEEFLSEKLERI 162
 DB 121 VRGSDPVKLKLDNGNIAEELSLIKWNTDSVEEFLSEKLERI 162

RESULT 2
 ID Q9ERR7 PRELIMINARY; PRT; 162 AA.
 AC Q9ERR7;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE 15 KDA SELENOPROTEIN.
 GN SEPI5.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20519655; PubMed=10945981;
 RA Kumaraswamy E., Malykh A., Korotkov K.V., Kozyavkin S., Hu Y.,
 RA Kwon S.Y., Moustafa M.E., Carlson B.A., Berry M.J., Lee B.D.,
 RA Hatfield D.L., Diamond A.M., Gladyshev V.N.;
 RT "Structure-Expression Relationships of the 15-kDa Selenoprotein Gene.
 RT POSSIBLE ROLE OF THE PROTEIN IN CANCER ETIOLOGY.";
 RT J. Biol. Chem. 275:35540-35547(2000).
 DR EMBL: AF288740; AAC31765.1; -.
 DR MGI:1927947; Sep15.
 SQ SEQUENCE 162 AA; 17738 MW; 6AB4B37D8C8B78447 CRC64;

Query Match 90.38; Score 750; DB 11; Length 162;
 Best Local Similarity 92.08; Pred. No. 2.5e-71;
 Matches 149; Conservative 1; Mismatches 12; Indels 0; Gaps 0;

QY 1 MAAGPSGCLVPAFGKRLILATVLAQVSAFGAEFSSSEACRELGFSNNLCCSCDILGQFNL 60
 DB 1 MAAGSGMLRPLALGRLILATVLAQVSAFGAEFSSSEACRELGFSNNLCCSCDILGQFNL 60
 QY 61 LQIDPDCGCCGCEAOFETKTLVAGAILLEVCGXKLGRRPVOAVRSDPKPLFRGLQIKY 120
 DB 61 LQIDPDCGCCGCEAOFETKTLVAGAILLEVCGXKLGRRPVOAVRSDPKPLFRGLQIKY 120
 QY 121 VRGSDPVKLKLDNGNIAEELSLIKWNTDSVEEFLSEKLERI 162
 DB 121 VRGSDPVKLKLDNGNIAEELSLIKWNTDSVEEFLSEKLERI 162

RESULT 3
 ID Q923V8 PRELIMINARY; PRT; 162 AA.
 AC Q923V8;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE 15 KDA CYTOSOLIC SELENOPROTEIN.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Boethlein D., Kyriakopoulos A., Behne D.;
 RT "A 15 kDa-selenoprotein in several tissues of the rat."
 RT Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF90544; AAK73100.1; -.
 SQ SEQUENCE 162 AA; 17706 MW; 65E939968CB872AC CRC64;

Query Match 89.58; Score 744; DB 11; Length 162;
 Best Local Similarity 92.68; Pred. No. 1.1e-70;
 Matches 150; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 1 MAAGPSGCLVPAFGKRLILATVLAQVSAFGAEFSSSEACRELGFSNNLCCSCDILGQFNL 60
 DB 1 MAAGSGMLRPLALGRLILATVLAQVSAFGAEFSSSEACRELGFSNNLCCSCDILGQFNL 60
 QY 61 LQIDPDCGCCGCEAOFETKTLVAGAILLEVCGXKLGRRPVOAVRSDPKPLFRGLQIKY 120
 DB 61 LQIDPDCGCCGCEAOFETKTLVAGAILLEVCGXKLGRRPVOAVRSDPKPLFRGLQIKY 120
 QY 121 VRGSDPVKLKLDNGNIAEELSLIKWNTDSVEEFLSEKLERI 162
 DB 121 VRGSDPVKLKLDNGNIAEELSLIKWNTDSVEEFLSEKLERI 162

RESULT 4
 ID Q9BS64 PRELIMINARY; PRT; 92 AA.
 AC Q9BS64;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE HYPOTHETICAL 9.7 KDA PROTEIN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=RENAL ADENOCARCINOMA;
 RA Strausberg R.;
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BONE MARROW, CHRONIC MYELOGENOUS LEUKEMIA;
 RA Strausberg R.;
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC005294; AAH05294.1; -.
 DR EMBL: BC016359; AAH16359.1; -.
 SQ SEQUENCE 92 AA; 9669 MW; 59EABAD5F7FAD58 CRC64;

Query Match 57.08; Score 474; DB 4; Length 92;
 Best Local Similarity 98.98; Pred. No. 1.7e-42;
 Matches 91; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAAGPSGCLVPAFGKRLILATVLAQVSAFGAEFSSSEACRELGFSNNLCCSCDILGQFNL 60
 DB 1 MAAGPSGCLVPAFGKRLILATVLAQVSAFGAEFSSSEACRELGFSNNLCCSCDILGQFNL 60
 QY 61 LQIDPDCGCCGCEAOFETKTLVAGAILLEVCG 92
 DB 61 LQIDPDCGCCGCEAOFETKTLVAGAILLEVCG 92

RESULT 5
 ID Q9VVJ7 PRELIMINARY; PRT; 142 AA.
 AC Q9VVJ7;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
 DE CG7484 PROTEIN.
 GN CG7484.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;


```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RX MEDLINE-20196006; PubMed-10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amaratilake P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zheng Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Butts K.C., Bussem D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cavley S., Dahlke C., Davernport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferliera S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jajaji M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Minkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pauley J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klimos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spreading A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weissstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri U.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003523; AAF49314.1; -
DR FlyBase: FBgn0036745; CG7484.
SQ SEQUENCE 142 AA; 15965 MW; 01D06725BE3DB4DF CRC64;

Query Match 37.4%; Score 311; DB 5; Length 142;
Best Local Similarity 51.8%; Pred. No. 4.6e-25;
Matches 59; Conservative 17; Mismatches 38; Indels 0; Gaps 0;

QY 48 LCSSCDLLGQFNLLQJDPDCRCCEAEAFETKRLKLYAGAILVEYCGKXKLRFPQVQAFVRS 107
DB 1 MCSSCEKLDLDFDITRIPQCKQCCTLDQPAQAFYAKAILVEYCGKCFRAPHYQIQAFIS 60
QY 108 DKRFLEQLQIVYVGSPPVYKLLDNGNIAELSLTKNNTSVEEFLSEKLER 161
DB 61 GRPAFEPNLIQIKYVGRGLDPVVKLLDASGRVQETLSITKWNITDVEEFETHLAK 114

RESULT 6
Q9N4C6 PRELIMINARY; PRT; 152 AA.
AC 09N4C6;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE HYPOTHEtical 17.6 KDA PROTEIN.
GN Y76B12C.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE-99069613; PubMed-9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Cordes M.;
RT "The sequence of C. elegans cosmid Y76B12C.";
RL Submitted (Mar-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC024211; AAF36064.2; -
KW Hypothetical protein.
SQ SEQUENCE 152 AA; 17593 MW; 9CAA2F9C0ABE5B87 CRC64;

Query Match 30.1%; Score 250.5; DB 5; Length 152;
Best Local Similarity 40.7%; Pred. No. 1.2e-18;
Matches 61; Conservative 26; Mismatches 54; Indels 9; Gaps 5;

QY 17 LLLATVLAQVSAFGA----EFSSEACRELGFS-SNLCSSCDLLGQFNLLQJDPDCRCGC 71
DB 6 LLLAAVAVSPV--FGEVEEKRIDVECKAGFENETKCGICERLSYHLETLLTDLQCC 63
QY 72 OEBAQEFKRLKLYAGAILVEYCGKXKLRFPQVQAFVRSQKPLFRG-LQIKVRSQDPVYKL 130
DB 64 IREKEERKHEK-YFALLEVCECLARFPQVQAVHDMARFGKTKVYKVRQYVAL 122
QY 131 LDDNGNIAELSLTKNNTSVEEFLSEKLE 160
DB 123 KDADEFKTEKLVSEKWDPTLLDFFNQMLE 152

RESULT 7
Q9STL3 PRELIMINARY; PRT; 158 AA.
AC 09STL3;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE F3F20.17 PROTEIN.
GN F3F20.17.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RA Altai H., Araujo R., Palm C.J., Conway A.B., Conn L., Hansen N.F.,
RA Gonzalez A., Kremenetskaia I., Kim C., Lenz C., Li J., Liu S.,
RA Lueros S., Schwartz J., Shinn P., Toriumi M., Vysotskaya V.S.,
RA Walker M., Yu G., Ecker J., Theologis A., Davis R.W.;
RL Submitted (May-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC007153; AAD30621.1; -
DR InterPro: IPR00345; CytC_heme_bind.
DR PROSITE: PS00190; CYTOCHROME_C; UNKNOMN.1.
SQ SEQUENCE 158 AA; 17901 MW; A181CB37DA4A33B8B CRC64;

Query Match 27.3%; Score 226.5; DB 10; Length 158;
Best Local Similarity 32.9%; Pred. No. 4.3e-16;

```

[illegible]

```

RESULT 10
O94594 PRELIMINARY; PRT; 653 AA.
AC O94594;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HEAT SHOCK PROTEIN 70.
GN HSP70.
OS Leishmania infantum.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5671;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-LEM75;
RX MEDLINE=97172530; PubMed=9020174;
RA Oujada L., Soto M., Alonso C., Requena J.M.;
RT *Analysis of post-transcriptional regulation operating on
RT transcription products of the tandemly linked Leishmania infantum
RT hsp70 genes.*;
RL J. Biol. Chem. 272:4493-4499(1997).
CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
DR EMBL; Y08020; CA69282.1; -.
DR HSSP; P08107; IHO.
DR InterPro; IPR001023; HSP70.
DR Pfam; PF00012; HSP70.1.
DR PRINTS; PR00301; HEATSHOCK70.
DR PROSITE; PS00297; HSP70_1; 1.
DR PROSITE; PS00329; HSP70_2; 1.
DR PROSITE; PS01036; HSP70_3; 1.
DR ATP-binding; Heat shock.
KW SEQUENCE 653 AA; 71078 MW; 908669ABB8B20BBF CRC64;
SQ

Query Match 9.2%; Score 76.5; DB 5; Length 653;
Best Local Similarity 27.9%; Pred. No. 16;
Matches 24; Conservative 14; Mismatches 35; Indels 13; Gaps 2;

OY 48 LCSSCDLLGQFNLDLPDPCGCCGCEAOFETKKLYAGAILLEVCCKLGRFPQVAFVRS 107
DB 451 MKKDCHLGTFDLGSGIPRPRGLPQIEVTFD---LDANGILNVAEKGCTGRKRNQITTN 507
OY 108 DKPKLFRGLQIKYVNGSDPYLKLLDD 133
DB 508 DKGRL-----SKDEIERMVND 523

RESULT 11
O25292 PRELIMINARY; PRT; 653 AA.
AC O25292;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HEAT-SHOCK PROTEIN (FRAGMENT).
GN HSP70.
OS Leishmania infantum.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5671;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-LEM75 /ZYMODEME 1;
RX MEDLINE=9635058; PubMed=8728991;
RA Oujada L., Requena J.M., Soto M., Alonso C.;
RT *During canine viscero-cutaneous leishmaniasis the anti-Hsp70
RT antibodies are specifically elicited by the parasite protein.*;
RL Parasitology 112:277-284(1996).
CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
DR EMBL; X85798; CA59793.1; -.
DR HSSP; P08107; IHO.

```

```

DR InterPro; IPR001023; HSP70.
DR Pfam; PF00012; HSP70.1.
DR PRINTS; PR00301; HEATSHOCK70.
DR PROSITE; PS00297; HSP70_1; 1.
DR PROSITE; PS00329; HSP70_2; 1.
DR PROSITE; PS01036; HSP70_3; 1.
DR ATP-binding.
KW NON_TER 653
FT SEQUENCE 653 AA; 71163 MW; A4EAB21C07F5E08E CRC64;
SQ

Query Match 9.2%; Score 76.5; DB 5; Length 653;
Best Local Similarity 27.9%; Pred. No. 16;
Matches 24; Conservative 14; Mismatches 35; Indels 13; Gaps 2;

OY 48 LCSSCDLLGQFNLDLPDPCGCCGCEAOFETKKLYAGAILLEVCCKLGRFPQVAFVRS 107
DB 451 MKKDCHLGTFDLGSGIPRPRGLPQIEVTFD---LDANGILNVAEKGCTGRKRNQITTN 507
OY 108 DKPKLFRGLQIKYVNGSDPYLKLLDD 133
DB 508 DKGRL-----SKDEIERMVND 523

RESULT 12
O39830 PRELIMINARY; PRT; 664 AA.
AC O39830;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE BIP ISOFORM A.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosid 1; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX NCBI_TaxID=3847;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CENTURY 84; TISSUE-LEAF;
RX MEDLINE=95218610; PubMed=776051;
RA Kalinski A., Rowley D.L., Loer D.S., Foley C., Buta G., Herman E.M.;
RT *Binding-protein expression is subject to temporal, developmental and
RT stress-induced regulation in terminally differentiated soybean
RT organs.*;
RL Planta 195:611-621(1995).
CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
DR EMBL; U08384; AA81956.1; -.
DR HSSP; P19120; 3HSC.
DR InterPro; IPR000886; ER_target.
DR InterPro; IPR001023; HSP70.
DR InterPro; IPR000671; Hydrgrn_uptake.
DR Pfam; PF00012; HSP70.1.
DR PRINTS; PR00301; HEATSHOCK70.
DR PROSITE; PS00014; ER_TARGET; UNKNOWN_1.
DR PROSITE; PS00297; HSP70_1; 1.
DR PROSITE; PS00329; HSP70_2; 1.
DR PROSITE; PS01036; HSP70_3; 1.
DR ATP-binding.
KW SEQUENCE 664 AA; 73336 MW; E1E2ABCB2086E3D6 CRC64;
SQ

Query Match 9.2%; Score 76.5; DB 10; Length 664;
Best Local Similarity 22.2%; Pred. No. 16;
Matches 45; Conservative 26; Mismatches 69; Indels 63; Gaps 8;

OY 5 PSCGIVAFGRKLLATVQ-----AVSAFGAESSEACRELGFSNLLCSCDILGQF 58
DB 444 PRVTVIPT--KRSQVFTTYQDOOSTVSIQVFEGERS-----LTKCKRLGRF 488
OY 59 NLLQDPDRCGCCGCEAOFETKKLYAGAILLEVCCKLGRFPQVAFVRS DKPKLFRGLQI 118
DB 489 ELGIPRPRGRGPQIEVTFVD---ANGILNVAEDKGTGKSEKITITNEKRLSOEIE 545

```

ID	Accession	Species	Score	DB	Length	Score	DB	Length
QY	119	KYVRCSD	-----PVKILND	1	311	141		
DB	546	KMVAEEFAEDDKVKVRIDARNSLEFYUYNMKNQVSDXDKLADKLSPDEKEKVEPTAK	605					
QY	142	SLKMWTD-SVEEFLSEKLERI	162					
DB	606	EALWLDNQNSVEKREYEKTEK	628					
RESULT	13							
P73692	P73692	PRELIMINARY	PRF	265	AA			
AC	P73692							
DT	01-FEB-1997	(TREMBLrel. 02, Created)						
DT	01-FEB-1997	(TREMBLrel. 02, Last sequence update)						
DT	01-JUN-2001	(TREMBLrel. 17, Last annotation update)						
DE		HYPOTHEETICAL 29.8 KDA PROTEIN.						
GN	SLR1809							
OS		Synechocystis sp. (strain PCC 6803).						
OC		Bacteria; Cyanobacteria; Chroococcales; Synechocystis.						
OX		NCBI_TaxID=1148;						
RN		[1]						
RP		SEQUENCE FROM N.A.						
RA		MEDLINE=97061201; PubMed=8905231;						
RA		Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,						
RA		Miyajima N., Hirosewa M., Suglura M., Sasamoto S., Kimura T.,						
RA		Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naito K., Okumura S.						
RA		Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,						
RA		Tanaka S.;						
RT		"Sequence analysis of the genome of the unicellular cyanobacterium						
RT		Synechocystis sp. strain PCC6803. II. Sequence determination of the						
RT		entire genome and assignment of potential protein-coding regions.";						
RL		DNA Res. 3:109-136(1996).						
DR		EMBL; D90908; BAA17739.1; -						
DR		InterPro; IPR001440; IPR.						
KW		Hypothetical protein; Complete proteome.						
SO		SEQUENCE 265 AA; 29760 MW; 3A344CF7BDC25569 CRC64;						
Query Match		9.18; Score 76; DB 16; Length 265;						
Best Local Similarity		23.28; Pred. NO. 6.1;						
Matches	41; Conservative	22; Mismatches	60; Indels	54; Gaps				
QY	31	AEEFSEACREGEFSSNLLCSSCDLLGOF	-----NLIDDPDRCGCC	73				
DB	74	AELDDAIAFNGEGSEFLSSMAALGKKNKONKRRADQDNHLYKRGSLSEPGSEGLD	133					
QY	74	EAQETTK	-----LYGAILVEYCGKL	-----GREPVQAFVRSDKRPLFRGIQ	117			
DB	134	LAEEFRROKRYFRFTVGSIMFAGAL	-CGGGIYLLSGGILP	-FVIEEKAL	ENLQ	186		
QY	118	IKVYVSGSPVLIKLDNDGNINAEELST	-----KKNTPSVSEF	-----LSRKLE	160			
DB	187	STIEAQNDFIYALMGDEALRERFDTLISQOLDENRLNGLENTLQRLALEEKIK	243					
RESULT	14							
Q94785	Q94785	PRELIMINARY	PRF	220	AA			
AC	Q94785							
DT	01-FEB-1997	(TREMBLrel. 02, Created)						
DT	01-JAN-1998	(TREMBLrel. 05, Last sequence update)						
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)						
DE		HEAT SHOCK PROTEIN 70 (FRAGMENT).						
OS		Trypanosoma cruzi.						
OC		Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.						
OX		NCBI_TaxID=5693;						
RN		[1]						
RP		SEQUENCE FROM N.A.						
RC		STRAIN Y.						
RA		Tanaka M., Tanaka T., Mitsui Y., Yamamoto M., Wood J.N.;						
RT		"The parasite genome project: Trypanosoma cruzi.";						

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2002, 16:30:50 ; Search time 24.1 Seconds

(without alignments)
260.272 Million cell updates/sec

Title: US-09-676-718A-1

Perfect score: 831
Sequence: 1 MAAGPSGCLVPAGFKRLLLA.....ILKWNDSVEEFLSEKLERI 162

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	821	98.8	162	SE15_HUMAN	060613 homo sapien
2	756	91.0	162	SE15_RAT	0922v8 rattus norv
3	749	90.1	162	SE15_MOUSE	09e1r7 mus musculu
4	82	9.9	485	STE_RHIME	P15189 rhizobium m
5	81	9.7	653	HS70_LEIDO	P15804 leishmania
6	81	9.7	680	HS70_TRYCR	P05456 trypanosoma
7	80.5	9.7	661	HS74_TRYBB	P11445 trypanosoma
8	80	9.6	269	THRG_CAUCR	09e146 callobacter
9	80	9.6	652	HS70_LEIMA	007437 leishmania
10	80	9.6	983	PTPN_RAT	063359 rattus norv
11	78	9.4	290	BIP1_TOBAC	003681 nicotiana t
12	78	9.4	292	BIP2_TOBAC	003682 nicotiana t
13	76.5	9.2	557	ILVD_BACSU	P51785 bacillus su
14	76	9.1	668	BIP5_TOBAC	003685 nicotiana t
15	74.5	9.0	516	HS70_LEIMA	P14834 leishmania
16	74.5	9.0	667	GR78_APLCA	016936 aplysia cal
17	74.5	9.0	670	GR78_YARLI	099170 yarrowia li
18	74.5	9.0	684	FBL1_CHICK	073775 gallus gall
19	74	8.9	264	THRG_XYFLA	09p195 xyellella fas
20	74	8.9	663	THRG_SCHPO	P36604 schizosacch
21	74	8.9	979	PTPN_MOUSE	060673 mus musculu
22	73	8.8	293	BIP8_TOBAC	003686 nicotiana t
23	73	8.8	1120	MDP_RICPR	005955 rickettsia
24	72.5	8.7	228	BIP4_TOBAC	P27894 leishmania
25	72.5	8.7	667	BIP7_TOBAC	003684 nicotiana t
26	72.5	8.7	830	DYN1_MAIZE	P39055 caenorhabdl
27	72	8.6	663	BIP2_MAIZE	P24067 zea mays (m
28	71.5	8.6	643	HS71_DROME	P02825 drosophila
29	71.5	8.6	656	HS7C_DROME	P29844 drosophila
30	71.5	8.6	668	BIP_SPIOI	042434 spinacia ol
31	71	8.5	890	KPCN_HUMAN	094806 homo sapien
32	71	8.5	979	PTPN_BOVIN	P56722 bos taurus
33	71	8.5	979	PTPN_HUMAN	Q16649 homo sapien

34	70.5	8.5	212	1	DSBA_BUCAT	P57505 buchnera ap
35	70.5	8.5	432	1	SYN_LACDE	P54262 lactobacill
36	70.5	8.5	643	1	HS71_SCHPO	Q10265 schizosacch
37	70.5	8.5	659	1	YHH5_YEAST	P38760 saccharomyc
38	69.5	8.4	640	1	HS74_ANOAL	P41827 anopheles a
39	69	8.3	265	1	THRG_PSEAE	Q916b4 pseudomonas
40	69	8.3	666	1	BIP_LYCES	P49118 lycopersico
41	69	8.3	668	1	BIP2_ARATH	Q39043 arabidopsis
42	69	8.3	669	1	BIP1_ARATH	Q91kr3 arabidopsis
43	69	8.3	677	1	SPOT_RHEIN	P43811 haemophilus
44	69	8.3	682	1	GR78_YEAST	P16474 saccharomyc
45	69	8.3	715	1	LCCL_LACLA	Q9c1b8 lactococcus

ALIGNMENTS

RESULT 1
ID SE15_HUMAN STANDARD: PRT; 162 AA.
AC 060613; Q9NR01; Q9BS64; Q9GZM0;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE 15 kDa selenoprotein precursor.
GN SEPI5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RN SEQUENCE FROM N.A., AND SEQUENCE OF 98-106; 123-127 AND 146-158.
RX MEDLINE-98204881; PubMed-9535873;
RA Gladyshev V.N., Jeang K.-T., Wootton J.C., Hatfield D.L.,
RT "A new human selenium-containing protein. Purification,
RT characterization, and cDNA sequence."
RL J. Biol. Chem. 273:8910-8915(1998).
RN [2]
RN SEQUENCE FROM N.A.
RX MEDLINE-20519655; PubMed-10945981;
RA Kumaraswamy E., Malykh A., Korotkov K.V., Kozayvkin S., Hu Y.,
RA Kwon S.Y., Moustafa M.E., Carlson B.A., Berry M.J., Lee B.J.,
RA Hatfield D.L., Diamond A.M., Gladyshev V.N.,
RT "Structure-expression relationships of the 15-kDa selenoprotein gene.
RT Possible role of the protein in cancer etiology."
RL J. Biol. Chem. 275:35540-35547(2000).
RN [3]
RN SEQUENCE FROM N.A.
RA Ryu M., Moon E.,
RT "The human 15-kDa selenoprotein gene: characterisation of the genomic
RT structure and functional analysis of the promoter."
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RN SEQUENCE FROM N.A.
RA Pearce A.,
RT Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RN SEQUENCE FROM N.A.
RC TISSUE=Bone marrow, and Kidney;
RP Strausberg R.,
RT Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBMIT: SEEMS TO BE EITHER OLIGOMERIC OR HETEROOOLIGOMERIC.
CC -1- PTM: THE N-TERMINUS IS BLOCKED.
CC -1- MASS SPECTROMETRY: MW=14870; METHOD-Electrospray; RANGE=7-152.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----

DR EMBL: AF051994; AAC15478.1; -
DR EMBL: AF288991; AAG31556.1; -
DR EMBL: AF288992; AAG31557.1; -
DR EMBL: AF267986; AAF78966.1; ALT_SEQ.
DR EMBL: AF267986; AAF78966.1; JOINED.
DR EMBL: AF267983; AAF78966.1; JOINED.
DR EMBL: AF267984; AAF78966.1; JOINED.
DR EMBL: AF267985; AAF78966.1; JOINED.
DR EMBL: AL121989; CAC04186.1; ALT_SEQ.
DR EMBL: BC005294; AAH05294.1; ALT_TERM.
DR EMBL: BC016359; AAH16359.1; ALT_TERM.
DR MIM: 606254; -
KW Selenium; selenocysteine; signal.
FT CHAIN 1 25 POTENTIAL.
FT SE_CYS 26 162 15 KDA SELENOPROTEIN.
SQ SEQUENCE 162 AA; 17743 MW; 463EACB423B4CDBE CRC64;

Query Match 98.88; Score 821; DB 1; Length 162;
Best Local Similarity 98.88; Pred. No. 3.9e-75;
Matches 160; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 MAAGPSGCLVAFGKRLTLATVLAQVSAFGAEFSEACRELGFSNNLTCSSCDLLGQFNL 60
DB 1 MAAGPSGCLVAFGKRLTLATVLAQVSAFGAEFSEACRELGFSNNLTCSSCDLLGQFNL 60
OY 61 LQLPDPCGCCGCEAEAFETKRLIYAGAILLEVCGXKIGRPPVOAFVRSKPKLFRGLQIKY 120
DB 61 LQLPDPCGCCGCEAEAFETKRLIYAGAILLEVCGXKIGRPPVOAFVRSKPKLFRGLQIKY 120
OY 121 VRGSDPVKLDDNGNIAEELSIKWNDSVEEFLSEKLERI 162
DB 121 VRGSDPVKLDDNGNIAEELSIKWNDSVEEFLSEKLERI 162

RESULT 2
SE15_RAT STANDARD; PRT; 162 AA.
AC Q923V8;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE 15 kDa selenoprotein precursor.
CN SEP15.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Roethlislein D., Kyriakopoulos A., Behne D.;
RT "A 15 kDa selenoprotein in several tissues of the rat."
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF390544; AAK73100.1; ALT_SEQ.
KW Selenium; selenocysteine; signal.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 162 15 KDA SELENOPROTEIN.
FT SE_CYS 93 93
SQ SEQUENCE 162 AA; 17775 MW; 73985E968C9F36FC CRC64;

Query Match 91.08; Score 756; DB 1; Length 162;
Best Local Similarity 92.68; Pred. No. 1.2e-68;

Matches 150; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

OY 1 MAAGPSGCLVAFGKRLTLATVLAQVSAFGAEFSEACRELGFSNNLTCSSCDLLGQFNL 60
DB 1 MAAGGCGWLRLPALGLRLLATVLAQVSAFGAEFSEACRELGFSNNLTCSSCDLLGQFNL 60
OY 61 LQLPDPCGCCGCEAEAFETKRLIYAGAILLEVCGXKIGRPPVOAFVRSKPKLFRGLQIKY 120
DB 61 LQLPDPCGCCGCEAEAFETKRLIYAGAILLEVCGXKIGRPPVOAFVRSKPKLFRGLQIKY 120
OY 121 VRGSDPVKLDDNGNIAEELSIKWNDSVEEFLSEKLERI 162
DB 121 VRGSDPVKLDDNGNIAEELSIKWNDSVEEFLSEKLERI 162

RESULT 3
SE15_MOUSE STANDARD; PRT; 162 AA.
ID SE15_MOUSE
AC O9ERR7;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE 15 kDa selenoprotein precursor.
CN SEP15.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=20519655; Pubmed=10945981;
RA Kumaraswamy E., Malykh A., Korotkov K.V., Kozyavkin S., Hu Y.,
RA Kwon S.Y., Moustafa M.E., Carlson B.A., Berry M.D., Lee B.D.,
RA Hatfield D.L., Diamond A.M., Gladyshev V.N.;
RT "Structure-expression relationships of the 15-kDa selenoprotein gene.
RT Possible role of the protein in cancer etiology."
RL J. Biol. Chem. 275:35540-35547(2000).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF288740; AAG31765.1; ALT_SEQ.
DR MGD: MGI:1927947; Sep15.
KW Selenium; selenocysteine; signal.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 162 15 KDA SELENOPROTEIN.
FT SE_CYS 93 93
SQ SEQUENCE 162 AA; 17731 MW; 771E047D8C9F3C17 CRC64;

Query Match 90.18; Score 749; DB 1; Length 162;
Best Local Similarity 91.48; Pred. No. 6.2e-68;
Matches 148; Conservative 1; Mismatches 13; Indels 0; Gaps 0;

OY 1 MAAGPSGCLVAFGKRLTLATVLAQVSAFGAEFSEACRELGFSNNLTCSSCDLLGQFNL 60
DB 1 MAAGGCGWLRLPALGLRLLATVLAQVSAFGAEFSEACRELGFSNNLTCSSCDLLGQFNL 60
OY 61 LQLPDPCGCCGCEAEAFETKRLIYAGAILLEVCGXKIGRPPVOAFVRSKPKLFRGLQIKY 120
DB 61 LQLPDPCGCCGCEAEAFETKRLIYAGAILLEVCGXKIGRPPVOAFVRSKPKLFRGLQIKY 120
OY 121 VRGSDPVKLDDNGNIAEELSIKWNDSVEEFLSEKLERI 162
DB 121 VRGSDPVKLDDNGNIAEELSIKWNDSVEEFLSEKLERI 162

RESULT 4

DT 01-FEB-1991 (Rel. 17, last annotation update)
 DN Heat shock 70 kDa protein.
 OS HSP70.
 OS Trypanosoma cruzi.
 OS Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
 NCBI_TaxID=5693;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=8157703; PubMed=2831499;
 RA Requena J.M., Lopez M.C., Jimenez-Ruiz A., de la Torre J.C.,
 Alonso C.;
 RT "A head-to-tail tandem organization of hsp70 genes in Trypanosoma cruzi."
 RL Nucleic Acids Res. 16:1393-1406(1988).
 CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X07083; CA30115.1; -
 DR PIR: A28531; A28531.
 DR PIR: S06158; S06158.
 DR HSP: P08107; 1HUO.
 DR InterPro: IPR001023; HSP70.
 DR Pfam: PF00012; HSP70; 1.
 DR PRINTS: PR00301; HEATSHOCK70.
 DR PROSITE: PS00297; HSP70_1; 1.
 DR PROSITE: PS00329; HSP70_2; 1.
 DR PROSITE: PS01036; HSP70_3; 1.
 DR ATP-binding; Heat shock; Multigene family.
 KW APP-binding; Heat shock; Multigene family.
 SQ SEQUENCE 680 AA; 73809 MW; D102000F5B7A8D82 CRC64;
 Query Match 9.7%; Score 81; DB 1; Length 680;
 Best Local Similarity 27.8%; Pred. No. 2;
 Matches 37; Conservative 16; Mismatches 52; Indels 28; Gaps 4;
 QY 48 LCSSCDLLGQFNLLQDPDRCGCCOEAFETKRLVAGAILVECGXKLRPPVOAFAVRS 107
 DB 452 MTKDCHLLGTGTELSGIPRPGVPIEYTFD---LDANGILVSAEKEGTGRNQIYITN 508
 QY 108 DKPKLFRGLQTKYVR-----GSDPYLKLDDNG-----NIAEEL--S 142
 DB 509 DKGRISRAEIRMYAEAKYAEEDKQYQIDAKNGLENYAFSMKNVNDPNVAGKIDDA 568
 QY 143 ILKWNDSVEEFL 155
 DB 569 DKTTTSVVEAL 581
 RESULT 7
 HS74_TRYBB STANDARD; PRT; 661 AA.
 AC P1145;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-NOV-1990 (Rel. 16, last sequence update)
 DT 01-MAY-1992 (Rel. 22, last annotation update)
 DE Heat shock 70 kDa protein 4 (HSP70).
 OS Trypanosoma brucei brucei.
 OS Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
 NCBI_TaxID=5702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87088827; PubMed=3796613;
 RA Glass D.J., Polyere R.I., van der Ploeg L.H.T.;
 RT "Conserved sequences and transcription of the hsp70 gene family in
 RT Trypanosoma brucei."
 RL Mol. Cell. Biol. 6:4657-4666(1986).

CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M14697; AAA30204.1; -
 DR PIR: A25398; A25398.
 DR HSP: P19120; 1HPW.
 DR InterPro: IPR001023; HSP70.
 DR Pfam: PF00012; HSP70; 1.
 DR PRINTS: PR00301; HEATSHOCK70.
 DR PROSITE: PS00297; HSP70_1; 1.
 DR PROSITE: PS00329; HSP70_2; 1.
 DR PROSITE: PS01036; HSP70_3; 1.
 KW ATP-binding; Heat shock; Multigene family.
 SQ SEQUENCE 661 AA; 71433 MW; F9F927003E595A13 CRC64;
 Query Match 9.7%; Score 80.5; DB 1; Length 661;
 Best Local Similarity 24.7%; Pred. No. 2.2;
 Matches 36; Conservative 16; Mismatches 55; Indels 39; Gaps 4;
 QY 48 LCSSCDLLGQFNLLQDPDRCGCCOEAFETKRLVAGAILVECGXKLRPPVOAFAVRS 107
 DB 452 MTKDCHLLGTGTELSGIPRPGVPIEYTFD---LDANGILVSAEKEGTGRNQIYITN 508
 QY 108 DKPKLFR-----GLQ-----IKYRGSDPYLKLDDN 134
 DB 509 DKGRISRAEIRMYSDAKYAEEDKAHYXIDAKNGLENYAFSMKNITNDPNVAGKIDDA 568
 QY 135 GNIAEELSI--LKWNDSVEEFLSE 157
 DB 569 DKNAVTVAVEALRLNDQASLIE 594
 RESULT 8
 THIG_CAUCR STANDARD; PRT; 269 AA.
 ID THIG_CAUCR
 AC Q9A746;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, last sequence update)
 DT 16-OCT-2001 (Rel. 40, last annotation update)
 DE Thiazole biosynthesis protein thig.
 GN THIG OR C01880.
 OS Caulobacter crescentus.
 OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
 OC Caulobacter.
 NCBI_TaxID=69394;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 19089 / CB15;
 RX MEDLINE=21173698; PubMed=11259647;
 RA Nierman W.C., Feldblum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
 RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
 RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
 RA Debay R.T., Dodson R.J., Durkin A.S., Gwin M.L., Haft D.H.,
 RA Kolony J.F., Smit J., Craven M.B., Khouiri H., Shetty J., Berry K.,
 RA Ullendack T., Iran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
 RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
 RT "Complete genome sequence of Caulobacter crescentus."
 RT Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
 RL -1- FUNCTION: Required for the synthesis of the thiazole moiety of
 CC thiamine (By similarity).
 CC -1- COFACTOR: FMN (Potential).
 CC -1- PATHWAY: Thiamine biosynthesis.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE THIG FAMILY.
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: AE005862; AAK23855.1; -
 DR TIGR: AC1880; -
 DR InterPro: IPR003009; FMN_enzyme.
 KW Thiamine biosynthesis; Flavoprotein; FMN; Complete proteome.
 SQ SEQUENCE 269 AA; 28707 MW; 96E557AC69BA922 CRC64;

Query Match 9.6%; Score 80; DB 1; Length 269;
 Best Local Similarity 29.4%; Pred. No. 0.87;
 Matches 42; Conservative 15; Mismatches 66; Indels 20; Gaps 5;

OY 9 LVPARGLLATVYIQAVSAFQAESESEACRELGSSNLLCSCDLL-----GQFNLIQ 62
 DB 29 LVTGKGYDYATNMAAARAAAGAEIVTAVRRV---NLTDSQPLLVYVKPTEFTYL- 83
 OY 63 LDPDRGCCQEAQFETKRLYAGA-----ILEYCGKRLGRFPQVQAFYRSQKPLFRGL 116
 DB 84 --PNTAGCCTGDAVFTLRLARAGGMDLVKLEVSDPTLPEDMETRLSLKLAVDGE 141
 OY 117 QIKYVRSQDPV-LKLIDNGNIA 138
 DB 142 QVMVYCSDDPVYARKLEAGAVA 164

RESULT 9
 HS70_LEIAM STANDARD: PRT; 652 AA.
 ID HS70_LEIAM
 AC 007437;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE Heat shock 70 kDa protein.
 GN HSP70.
 OS Leishmania amazonensis.
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 OX NCBI_TaxID=5659;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MHOM/BR/77/LTB0016/C1S1;
 RX MEDLINE=94187793; PubMed=8139614;
 RA Bock J.H., Langer P.J.,
 RT "Sequence and genomic organization of the hsp70 genes of Leishmania
 RT amazonensis.";
 RL MOL. BIOCHEM. PARASITOL. 62:187-198(1993).
 CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: L14604; AAA53690.1; -
 DR EMBL: L14605; AAD15233.1; -
 DR HSSP: P08107; IHJO.
 DR InterPro: IPR001023; HSP70.
 DR Pfam: PF00012; HSP70.1.
 DR PRINTS: PR00301; HEATSHOCK70.
 DR PROSITE: PS00297; HSP70_1; 1.
 DR PROSITE: PS00329; HSP70_2; 1.
 DR PROSITE: PS01036; HSP70_3; 1.
 KW ATP-binding; Heat shock; Multigene family.
 FT VARIANT 81 81 L->S.

FT VARIANT 632 632 S -> G.
 SQ SEQUENCE 652 AA; 71197 MW; 0B21375DC05FE633 CRC64;

Query Match 9.6%; Score 80; DB 1; Length 652;
 Best Local Similarity 24.0%; Pred. No. 2.4;
 Matches 37; Conservative 17; Mismatches 58; Indels 42; Gaps 4;

OY 48 LCSSCDLLQGFNLQDPPDRGCCQEAQFETKRLYAGAILVEYCGKRLGRFPQVQAFVRS 107
 DB 452 MTRKDCILGTFDLGSGIPAPRGVQIEVFD---LDANGILNYSAEKGTGRNQITITN 508
 OY 108 DKPKLR-----GLQ-----IKYVRSQDPVYKLLD- 133
 DB 509 DKGRLSKDEIERVNDAMKYEADDKAQAADVREAKNGLENYAYSMKWTLGDSNVSGRLDPT 568
 OY 134 -----NGNTAEELSTLKMNDSVEEFLSEKLEKI 162
 DB 569 DKSTLNKEIEALEWISSNOEATKEEYEHKQEL 602

RESULT 10
 PRPN_RAT STANDARD: PRT; 983 AA.
 ID PRPN_RAT
 AC 063259; Q62883; Q63795; Q64643;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Protein-tyrosine phosphatase-like N precursor (R-PRP-N) (105 kDa 1st let
 DE cell antigen) (ICA105) (PRP IR-2) (PTLP) (Brain-enriched membrane-
 DE associated protein tyrosine phosphatase) (BEM-3).
 GN PRPN.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96016179; PubMed=7568143;
 RA Passini N., Larigan J.D., Genovese S., Appella E., Sinsaglla F.,
 RA Rogge L.;
 RT "The 37/40-kilodalton autoantigen in insulin-dependent diabetes
 RT mellitus is the putative tyrosine phosphatase IA-2.";
 RT Proc. Natl. Acad. Sci. U.S.A. 92:9412-9416(1995).
 RN [2]
 RP SEQUENCE OF 13-983 FROM N.A.
 RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Brain;
 RX MEDLINE=95386708; PubMed=7657822;
 RA Payton M.A., Hawkes C.J., Christie M.R.;
 RT "Relationship of the 37,000- and 40,000-M(r) tryptic fragments of
 RT I-slet antigens in insulin-dependent diabetes to the protein tyrosine
 RT phosphatase-like molecule IA-2 (ICA512).";
 RL J. Clin. Invest. 96:1506-1511(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95194305; PubMed=7887886;
 RA Kamayashi Y., Takahashi K., Bardhan S., Inagami T.;
 RT "Cloning and expression of protein tyrosine phosphatase-like protein
 RT derived from a rat pheochromocytoma cell line.";
 RL Biochem. J. 306:331-335(1995).
 RN [4]
 RP SEQUENCE OF 396-983 FROM N.A.
 RC STRAIN=Wistar; TISSUE=Brain;
 RA Itoh S., Okada M., Nakagawa H.;
 RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: IMPLICATED IN NEUROENDOCRINE SECRETORY PROCESSES. MAY BE
 CC INVOLVED IN PROCESSES SPECIFIC FOR NEUROSECRETORY GRANULES, SUCH
 CC AS THEIR BIOGENESIS, TRAFFICKING OR REGULATED EXOCYTOSIS OR MAY
 CC HAVE A GENERAL ROLE IN NEUROENDOCRINE FUNCTIONS. SEEMS TO LACK
 CC INTRINSIC ENZYME ACTIVITY (BY SIMILARITY).
 CC -1- FUNCTION: MAY FUNCTION AS A NEGATIVE REGULATOR OF PTASES IN
 CC NEURONAL TISSUES.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.


```

DB 71 PRNTVPT--KKSQVFTTQDOOTVSIQVEEGRS-----LTKDCRLGKF 115
QY 59 NLLQDLPDRCGCCOEAEQFETKLYAGALILEVCGKXKLGRRPOVQAFVRSKPKLFRGLQI 118
DB 116 DLGTGAPARGPQIETFEVD---ANGILNVAEDKGKSGKSKITITNDKGL-----166
QY 119 KYVKGSDPVKLKLDNGNIAEELSLKMNND---SVEEF-----LSEKLE 160
DB 167 ----SQEIERMVRFAEAEFAEDKKYKERIDANSLIETTYVNNKNOINDKDLADKLE 220

```

RESULT 12

```

BIP2_TOBAC STANDARD; PRT; 292 AA.
ID BIP2_TOBAC STANDARD; PRT; 292 AA.
AC 003682;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Lumenal binding protein 2 (BIP 2) (78 kDa glucose-regulated protein
GN homolog 2) (GRP 78-2) (Fragment).
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; eusterids I; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92361242; PubMed=1822990;
RA Denoecke J., Goldman M.H., Demolder J., Seurinck J., Botterman J.;
RT "The tobacco lumenal binding protein is encoded by a multigene
RL plant Cell 3:1025-1035(1991).
RN [2]
RP ERRATUM.
RA Denoecke J., Goldman M.H., Demolder J., Seurinck J., Botterman J.;
RL Plant Cell 3:1251-1251(1991).
CC -1- FUNCTION: PROBABLY PLAYS A ROLE IN FACILITATING THE ASSEMBLY OF
CC MULTIMERIC PROTEIN COMPLEXES INSIDE THE ER.
CC -1- SUBCELLULAR LOCATION: Endoplasmic reticulum lumen.
CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL: X60059; CAA42661.1; -
DR PIR: S21878; S21878.
DR PIR: P00262; P00262.
DR HSPD, P19120; INCG.
DR InterPro: IPR000886; ER_target.
DR InterPro: IPR001023; HSP70.
DR Pfam: PF00012; HSP70.1.
DR PROSITE: PS00297; HSP70.1; PARTIAL.
DR PROSITE: PS00329; HSP70.2; PARTIAL.
DR PROSITE: PS01036; HSP70.3; PARTIAL.
DR PROSITE: PS00014; ER_TARGET; 1.
KW ATP-binding; Endoplasmic reticulum; Multigene family.
FT CARBOHYD 241 241 N-LINKED (GLCNAC... ) (POTENTIAL).
FT SITE 289 292 PREVENT SECRETION FROM ER (POTENTIAL).
FT SEQUENCE 292 AA: 32260 MW: 667A4F5107C6E2D5 CRC64;
SQ

```

Query Match 9.4%; Score 78; DB 1; Length 292;
 Best Local Similarity 23.0%; Pred.No.1.5;
 Matches 41; Conservative 24; Mismatches 63; Indels 50; Gaps 7;

```

QY 5 PSCGLVPARGKRLLATVLD-----AVSAFAGFEPSSEACRELGFSNNLICSSCDLLGOF 58
DB 71 PRNTVPT--KKSQVFTTQDOOTVSIQVEEGRS-----LTKDCRLGKF 115
QY 59 NLLQDLPDRCGCCOEAEQFETKLYAGALILEVCGKXKLGRRPOVQAFVRSKPKLFRGLQI 118
DB 116 DLGTGAPARGPQIETFEVD---ANGILNVAEDKGKSGKSKITITNDKGL-----166
QY 119 KYVKGSDPVKLKLDNGNIAEELSLKMNND---SVEEF-----LSEKLE 160
DB 167 ----SQEIERMVRFAEAEFAEDKKYKERIDANSLIETTYVNNKNOINDKDLADKLE 220

```

RESULT 13

```

ILVD_BACSU STANDARD; PRT; 557 AA.
ID ILVD_BACSU STANDARD; PRT; 557 AA.
AC P51785;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Dihydroxy-acid dehydratase (EC 4.2.1.9) (DAD) (vegetative protein
GN 110) (VEG110).
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / MARBURG;
RX MEDLINE=96349105; PubMed=8760912;
RA Sorokin A.V., Azevedo V., Zumbstein E., Galleron N., Ehrlich S.D.,
RA Serrro P.;
RT "Sequence analysis of the Bacillus subtilis chromosome region between
RT the sera and kdg loci cloned in a yeast artificial chromosome.";
RL Microbiology 142:2003-2016(1996).
RN [2]
RP SEQUENCE OF 1-15.
RC STRAIN=IS58;
RX MEDLINE=97443988; PubMed=9298659;
RA Antelmann H., Bernhard J., Schmid R., Mach H., Voelker U.,
RA Hecker M.;
RT "First steps from a two-dimensional protein index towards a response-
RT regulation map for Bacillus subtilis.";
RL Electrophoresis 18:1451-1463(1997).
CC -1- CATALYTIC ACTIVITY: 2,3-dihydroxy-3-methylbutanoate -> 3-methyl-2-
CC oxobutanoate + H(2)O.
CC -1- COFACTOR: BINDS 1 4FE-4S CLUSTER (POTENTIAL).
CC -1- PATHWAY: FOURTH STEP IN VALINE AND ISOLEUCINE BIOSYNTHESIS.
CC -1- SIMILARITY: BELONGS TO THE ILVD / EDD FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL: L77246; AAA96629.1; -
DR EMBL: 299115; CAB14105.1; -
DR Subtilist; BG11532; ILVD.
DR InterPro: IPR000581; ILVD_EDD.
DR Pfam: PF0002691; ILVD_EDD; 1.
DR ProDom: PD002691; ILVD_EDD; 1.
DR PROSITE: PS00886; ILVD_EDD_1; 1.
DR PROSITE: PS00887; ILVD_EDD_2; 1.
KW Branched-chain amino acid biosynthesis; Lyase; Iron-sulfur; 4Fe-4S;
KW complete proteome.
FT INIT MET 0
FT METAL 121 121 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 194 194 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT SEQUENCE 557 AA: 59417 MW: CAC59717EB9D7EDA CRC64;
SQ

```

Query Match 9.28; Score 76.5; DB 1; Length 557;
 Best Local Similarity 21.5%; Pred. No. 4.5;
 Matches 37; Conservative 33; Mismatches 75; Indels 27; Gaps 7;

1 MAAGPSCVPAPEKRLILATVLAQVSAFGA-ESSSEACRELGFSSNLLSCDILGQFN 59
 142 VSGGPMAGRTSGRKISLSVFGVAYQAGKINENLELEQFGCPGSGS--GMFT 199
 60 LLADPDCRGCCOEAEQETKLYAGAILLEVCGXKLGFRPQVAEVRSDPKPLFRGLQIK 119
 200 ANSMN-----CSEA-----LGLALPONGITLATSPEKKEFKVKSAAQIMETIR-K 244
 120 YVSGSDPV-LKLDL-----NGNTAEELSLIKMNTDVEEFLSEKLERI 162
 245 DIKPRDVTVAIDNALDMALGSGTNTVLTALANAGVEYSLERINEV 296

RESULT 14
 BIP5_TOBAC STANDARD; PRT; 668 AA.

ID BIP5_TOBAC STANDARD; PRT; 668 AA.
 AC 003685;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DE 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Luminal binding protein 5 precursor (Bip 5) (78 kDa glucose-regulated protein homolog 5) (GRP 78-5).
 GN BIP5.
 OS Nicotiana tabacum (Common tobacco).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; easterids I; Solanales; Solanaceae; Nicotiana.
 OX NCBI_TaxID=4097;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92361242; PubMed=1822990;
 RA Denecke J., Goldman M.H., Demolder J., Seurinck J., Botteman J.;
 RT "The tobacco luminal binding protein is encoded by a multigene family."
 RL Plant Cell 3:1025-1035(1991).
 RN [2]
 RP ERRATUM.
 RA Denecke J., Goldman M.H., Demolder J., Seurinck J., Botteman J.;
 RL Plant Cell 3:1251-1251(1991).
 CC -1- FUNCTION: PROBABLY PLAYS A ROLE IN FACILITATING THE ASSEMBLY OF MULTIMERIC PROTEIN COMPLEXES INSIDE THE ER.
 CC -1- SUBCELLULAR LOCATION: Endoplasmic reticulum lumen.
 CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X60058; CAA42660.1; -
 DR EMBL: X60058; S21880.
 DR PIR: J01361; J01361.
 DR HSSP: P19120; 3HSC.
 DR InterPro: IPR000886; ER_target.
 DR InterPro: IPR001023; HSP70.
 DR Pfam: PF00012; HSP70; 1.
 DR PRINTS: PR00301; HEATSHOCK70.
 DR PROSITE: PS00014; ER_TARGET; 1.
 DR PROSITE: PS00297; HSP70_1; 1.
 DR PROSITE: PS00329; HSP70_2; 1.
 DR PROSITE: PS01036; HSP70_3; 1.
 KM APP-binding; Endoplasmic reticulum; Multigene family; Signal.
 FT SIGNAL 1 24
 FT CHAIN 25 668 LUMINAL BINDING PROTEIN 5.

FT CARBOHYD 617 617 N-LINKED (GLCNAC...) (POTENTIAL).
 FT SITE 665 668 PREVENT SECRETION FROM ER (POTENTIAL).
 SQ SEQUENCE 668 AA; 73744 MW; 7980231A91DC590 CRC64;

Query Match 9.18; Score 76; DB 1; Length 668;
 Best Local Similarity 23.08; Pred. No. 6.2;
 Matches 41; Conservative 23; Mismatches 64; Indels 50; Gaps 7;

5 PSCGLVPAFGKRLILATVLAQ-----AVSAFGAEFSSEACRELGFSSNLLSCDILGQF 58
 447 PRNVIVPT--KSGVFTTYDQQTIVTISVFEGRS-----LTKCRILGKF 491
 59 NLADPDCRGCCOEAEQETKLYAGAILLEVCGXKLGFRPQVAEVRSDPKPLFRGLQI 118
 492 DLTIAPAPRGTPQIEVTFEVD--ANGIINVKAEKDRASGSEKRTITNDKGRF----- 542
 119 KYVSGSDPVLLKLDNNGINIAEELSLIKMNTD-----SVEEF-----LSEKLE 160
 543 -----SOEIERMVAEAEFEAEDEKKVKERIDARNSLETIYYNMKNQINDXDKLADKLE 596

RESULT 15
 HS70_LEIMA STANDARD; PRT; 516 AA.

ID HS70_LEIMA STANDARD; PRT; 516 AA.
 AC P14834;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Heat shock 70 kDa protein (Fragment).
 GN HSP70.
 OS Leishmania major.
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 OX NCBI_TaxID=5664;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=WR300.
 RX MEDLINE=89041551; PubMed=3186441;
 RA Lee M.G., Atkinson B.L., Gianini S.H., van der Ploeg L.H.T.;
 RT "Structure and expression of the hsp 70 gene family of Leishmania major."
 RL Nucleic Acids Res. 16:9567-9585(1988).
 CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M36675; AAA29251.1; -
 DR EMBL: M36675; S06443.
 DR PIR: S06443; S06443.
 DR HSSP: P08109; 1CKR.
 DR InterPro: IPR001023; HSP70.
 DR Pfam: PF00012; HSP70; 1.
 DR PROSITE: PS00297; HSP70_1; 1.
 DR PROSITE: PS00329; HSP70_2; 1.
 DR PROSITE: PS01036; HSP70_3; 1.
 KM APP-binding; Heat shock; Multigene family.
 FT NON_TER 516 516
 FT SEQUENCE 516 AA; 56535 MW; 36F83ACE952452 CRC64;

Query Match 9.08; Score 74.5; DB 1; Length 516;
 Best Local Similarity 33.88; Pred. No. 6.5;
 Matches 22; Conservative 9; Mismatches 31; Indels 3; Gaps 1;

48 LCSSCDLGGFNLADLPDCRGCCOEAEQETKLYAGAILLEVCGXKLGFRPQVAEVRSD 107
 452 MTKCHILGTDLGSGIRPARGVQIEVTD---LDANGILNVAEKEKGRKRNQITITN 508

OY 108 DKPKL 112
11:1
Db 509 DKGRL 513

Search completed: September 18, 2002, 16:36:05
Job time: 315 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2002, 15:28:40 ; Search time 23.92 Seconds

(without alignments)
165.424 Million cell updates/sec

Title: US-09-676-718a-1

Perfect score: 831
Sequence: 1 MAAGPSCGLVPARGKRLLA.....ILKWNDSVEEFLSEKLERI 162

Scoring table: BLOSUM62
Gapop 10.0 , Capext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents, AA: *
1: /cgn2_6/ptodata/2/1aa/5A.CONB.pep: *
2: /cgn2_6/ptodata/2/1aa/5B.CONB.pep: *
3: /cgn2_6/ptodata/2/1aa/6A.CONB.pep: *
4: /cgn2_6/ptodata/2/1aa/6B.CONB.pep: *
5: /cgn2_6/ptodata/2/1aa/PCRTUS.CONB.pep: *
6: /cgn2_6/ptodata/2/1aa/Backfile1.pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	79.5	9.6	661	1	US-08-399-986B-5
2	79.5	9.6	661	2	US-08-493-754A-5
3	74	8.9	663	1	US-08-441-139-7
4	69	8.3	682	1	US-08-441-139-2
5	66.5	8.0	753	4	US-08-942-686-2
6	65.5	7.9	187	6	5196523-13
7	65.5	7.9	654	1	US-08-441-139-11
8	65.5	7.9	666	1	US-08-441-139-16
9	65	7.8	1072	4	US-09-357-251-31
10	64.5	7.8	83	4	US-09-442-631-2
11	64.5	7.8	461	4	US-09-352-990-27
12	63.5	7.6	127	2	US-08-162-146-3
13	63.5	7.6	127	4	US-09-314-127-3
14	63.5	7.6	718	1	US-08-444-792-4
15	63.5	7.6	718	1	US-08-445-042-4
16	63.5	7.6	1454	4	US-08-392-459-26
17	63.5	7.6	1454	5	PCT-US91-08525-26
18	63.5	7.6	1454	5	PCT-US93-04384-8
19	63	7.6	2710	2	US-08-568-459A-12
20	63	7.6	2710	2	US-08-487-826B-12
21	63	7.6	3060	2	US-08-487-826B-14
22	62.5	7.5	187	3	US-08-912-768-3
23	62.5	7.5	366	2	US-08-984-171-4
24	62.5	7.5	456	2	US-08-910-731-2
25	62.5	7.5	456	2	US-08-795-395-2
26	62.5	7.5	768	1	US-08-454-455-4
27	62	7.5	372	2	US-08-846-762-79

28	62	7.5	423	1	US-08-484-324-2	Sequence 2, Appl1
29	62	7.5	423	1	US-08-486-409-2	Sequence 2, Appl1
30	62	7.5	423	2	US-08-825-784-2	Sequence 2, Appl1
31	62	7.5	423	2	US-08-972-258-2	Sequence 2, Appl1
32	62	7.5	423	4	US-09-263-128-2	Sequence 2, Appl1
33	62	7.5	423	5	PCT-US92-01899-2	Sequence 2, Appl1
34	62	7.5	773	3	US-08-564-264-1	Sequence 1, Appl1
35	62	7.5	781	4	US-09-193-503B-3	Sequence 3, Appl1
36	62	7.5	781	4	US-09-193-503B-6	Sequence 6, Appl1
37	62	7.5	781	4	US-09-193-503B-9	Sequence 9, Appl1
38	62	7.5	781	4	US-09-193-503B-11	Sequence 11, Appl1
39	61.5	7.4	236	2	US-08-494-907-8	Sequence 8, Appl1
40	61.5	7.4	236	3	PCT-US96-10986-8	Sequence 8, Appl1
41	61.5	7.4	272	1	US-08-698-551-8	Sequence 8, Appl1
42	61.5	7.4	272	2	US-08-602-228-8	Sequence 8, Appl1
43	61.5	7.4	272	2	US-08-649-341A-8	Sequence 8, Appl1
44	61.5	7.4	272	2	US-08-494-440B-8	Sequence 8, Appl1
45	61.5	7.4	272	2	US-08-533-901B-8	Sequence 8, Appl1

ALIGNMENTS

```
RESULT 1
US-08-399-986B-5
; Sequence 5, Application US/08399986B
; Patent No. 5801041
; GENERAL INFORMATION:
; APPLICANT: Godwin, Andrew K.
; TITLE OF INVENTION: NO. 5801041el Gene Associated with Suppression
; NUMBER OF INVENTIONS: Of Tumor Development
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/399,986B
; FILING DATE: 06-MAR-1995
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Hagan, Patrick J.
; REGISTRATION NUMBER: 27,643
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 661 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-399-986B-5

Query Match 9.6%; Score 79.5; DB 1; Length 661;
Best Local Similarity 20.3%; Pred. No. 0.35;
Matches 59; Conservative 28; Mismatches 64; Indels 139; Gaps 11;

0Y 8 CLVP-----AFGRLLATVIAV-----SAGCAFEFS-----EACRELGRS 44
DB 89 CLVPIONTGICIAMLYVEGRKLVVSTVQFIPSLQTLRTFTFNKDDSSIRIDIPQCKPLSPG 148
```

```

QY      45 SNLSCSDL-----LGFNL-----LDPPDRCGCGE-----73
           || :||
Db      149 EVLGCTSPRLDAASKDAIVYLGDGRFHLESTIMINPELEAFQDPYRKRTFREYYDHDLM 208
           |||
QY      74 -----ENQFET--KRLTYAGATLEVCSXKL 95
           |||
Db      209 RKNRGISIEARKCCTTGGTLIGTLRGOONLKVBELEAQERKKCKRTLRYLLSITPEKL 268
           |||
QY      96 GRFPQYVAFFVSIDPKL-----PGLQIQIKKYRGSDPV 127
           || :||
Db      269 AMFEPEVDWQVAVACPRLSIDMGTFPRPLLYPFELALVALDNVSPKRFCLQI---TGWTI 325
           |||
QY      128 LKLL-----DNGNIAEELSILIKNTDSVEEFSEKIERI 162
           |||
Db      326 IRMLPWVLGIIRIMKRYVRANGNGNLI--LLSKPKTHSRBELSYNEEKAKRI 374

```

RESULT 2
 US-08-493-754A-5
 Sequence 5, Application US/08493754A
 Patent No. 5821338
 GENERAL INFORMATION:
 APPLICANT: Godwin, Andrew K.
 TITLE OF INVENTION: No. 5821338el Gene Associated with Suppression
 TITLE OF INVENTION: of Tumor Development
 NUMBER OF SEQUENCES: 35
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Dann, Dorfman, Herrell and Skillman
 STREET: 1601 Market Street
 CITY: Philadelphia
 STATE: PA
 COUNTRY: USA
 ZIP: 19103-2307
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/493,754A
 FILING DATE: 22-JUN-1995
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Hagan, Patrick J.
 REGISTRATION NUMBER: 27,643
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (215) 563-4100
 TELEFAX: (215) 563-4044
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 661 amino acids
 TYPE: amino acid
 STRANDEDNESS: not relevant
 TOPOLOGY: not relevant
 MOLECULE TYPE: protein
 HYPOTHEICAL: NO
 ANTI-SENSE: NO
 US-08-493-754A-5

[illegible]

```

QY 74 -----EAGET--KRYAGALIEVCXKL 95
Db 209 RKNRGISIAKCKTFTGLIGTLGRGNLKVYBELAQLERGKRLARVLSEIPEKL 2686
QY 96 GRFOYQAIFYSDMKPKI-----FGLDIKVRGSDPY 127
Db 269 AMFEYDCHVYACPRKLSIDMGTFPRKPLPYFELVALDNVSKFKCLOD---TGQWTI 3255
QY 128 LKLL-----DDNGINAEBLSTLKWNTDSVEEPLSEKTERI 162
Db 326 IRMLPYWLGIRIIMKRYVANGNGNLI--LLSKPKIHSBELSTFPEEKKRI 374

```

```

RESULT 3
US-08-441-139-7
Sequence 7, Application US/08441139
Patent No. 5773245
GENERAL INFORMATION:
APPLICANT: Wiltup, Dr. Karl D.
APPLICANT: Robinson, Anne S.
TITLE OF INVENTION: METHODS FOR INCREASING SECRETION OF
TITLE OF INVENTION: RECOMBINANTLY EXPRESSED PROTEINS
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: NY
COUNTRY: USA
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/441,139
FILING DATE: 15-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/089,997
FILING DATE: 06-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: Digiglio, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 8646
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-742-4343
TELEFAX: 516-742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 663 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-441-139-7

```

[illegible]

RESULT 4

US-08-162-146-3

Query Match 7.6%; Score 63.5; DB 2; Length 127;
Best Local Similarity 26.7%; Pred. No. 3.3;
Matches 39; Conservative 13; Mismatches 45; Indels 49; Gaps 9;

QY 8 CLVP-AFGKRLATVIAVSAFAEFSSA---CRELGSSNLLGSSCDLLGQFNLLQL 63
DB 3 CLISLAG---LLA--LEAALALAPKFIAPVQVWCPEPSSSEETLCLIS----- 45
QY 64 DPDCRG---CCOEBAQETKLVAGALLEVCGXKLGREFPOVQAFV----- 105
DB 46 DNDCLGTVCCPSAAGGSCR-----TPITVPTKAGCHCPVQAPMLSQLCEELSDCANDI 100
QY 106 --RSDKPLFRGLQIKYVSGSDPYLK 129
DB 101 ECRGDKKCCFRCAMRYL---EPITL 123

RESULT 13

US-09-314-127-3
; Sequence 3, Application US/09314127
; Patent No. 6268545
; GENERAL INFORMATION:
; APPLICANT: HODDERLINE, Louis-Marie
; APPLICANT: DEVINOT, Eve
; APPLICANT: THEPOT, Dominique
; TITLE OF INVENTION: Production of a Protein of Interest in
; TITLE OF INVENTION: the Milk of a Transgenic Mammalian
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/314,127
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/162,146
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 91/07179
; FILING DATE: 12-JUN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: WEGNER, Harold C.
; REGISTRATION NUMBER: 25,258
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 127 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-314-127-3

Query Match 7.6%; Score 63.5; DB 4; Length 127;
Best Local Similarity 26.7%; Pred. No. 3.3;
Matches 39; Conservative 13; Mismatches 45; Indels 49; Gaps 9;

QY 8 CLVP-AFGKRLATVIAVSAFAEFSSA---CRELGSSNLLGSSCDLLGQFNLLQL 63
DB 3 CLISLAG---LLA--LEAALALAPKFIAPVQVWCPEPSSSEETLCLIS----- 45
QY 64 DPDCRG---CCOEBAQETKLVAGALLEVCGXKLGREFPOVQAFV----- 105
DB 46 DNDCLGTVCCPSAAGGSCR-----TPITVPTKAGCHCPVQAPMLSQLCEELSDCANDI 100
QY 106 --RSDKPLFRGLQIKYVSGSDPYLK 129
DB 101 ECRGDKKCCFRCAMRYL---EPITL 123

RESULT 14

US-08-444-792-4
; Sequence 4, Application US/08444792
; Patent No. 5726037
; GENERAL INFORMATION:
; APPLICANT: Bodary, Sarah C.
; APPLICANT: Gorman, Cornelia M.
; APPLICANT: McLean, John W.
; APPLICANT: Napier, Mary A.
; TITLE OF INVENTION: METHOD FOR PREPARING WATER SOLUBLE
; TITLE OF INVENTION: POLYPEPTIDES
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 720 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Winpatin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/444,792
; FILING DATE: 19-May-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/380227
; FILING DATE: 30-JAN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/218878
; FILING DATE: 28-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/821337
; FILING DATE: 13-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/444490
; FILING DATE: 01-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/290224
; FILING DATE: 22-DEC-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Lee, Wendy M.
; REGISTRATION NUMBER: 00,000
; REFERENCE/DOCKET NUMBER: P0552PIC3D4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/952-9881
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 718 amino acids
; TYPE: Amino Acid
; TOPOLOGY: linear
; US-08-444-792-4

Query Match 7.6%; Score 63.5; DB 1; Length 718;

Best Local Similarity 26.5%; Pred. No. 38;
Matches 35; Conservative 15; Mismatches 61; Indels 21; Gaps 7;

```
0y      16 RLLATIVQ--AVSFGFEFSEACRELGFSSNLTGSCDILGLGFNLGLDPGRGCCQ 73
        | | | | | : | : | | | | | | | | | | | | | | | | | | | |
Db      8 RPLMTVTLALCALGLGVGC-CPNICITTRGVSS---CQC-----LAASPMCAWCSDE 55
                                                | | | | | | | | | | | | | |
0y      74 EAQETTK--LYAGAILEVCXKLGREFPOVAFYRSDPKLFRG---LDIKYRGSDPV 127
        : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      56 ALPLGSPCDLKENLLKDNCAPESTIEFPVSFARVLEDRLPSDKSGDSOVTOVSPORIA 115
        : : : : : : : : : : : : : : : : : : : : : : : : : :
0y      128 LKL-L-DDGNIA 138
        | : | | | : | :
Db      116 LRLRPDDSKNF 127
```

RESULT 15
US-08-445-042-4

GENERAL INFORMATION:
APPLICANT: Bodary, Sarah C.
APPLICANT: Gorman, Cornelia M.
APPLICANT: McLean, John W.
APPLICANT: Napier, Mary A.
TITLE OF INVENTION: METHOD FOR PREPARING WATER SOLUBLE
POLYPEPTIDES
TITLE OF INVENTION: POLYPEPTIDES
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
000000

```

;      COMPUTER READABLE FORM:
;      MEDIUM TYPE: 3.5 inch, 720 Kb floppy disk
;

```

```

? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: WinPalin (Genetech)
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/445,042
? FILING DATE: 19-May-1995
? CLASSIFICATION: 530
?

```

Query Match	7.68;	Score 63.5;	DB 1;	Length 718;
Best Local Similarity	26.58;	Pred. No. 38;		
Matches 35;	Conservative 15;	Mismatches 61;	Indels 21;	Gaps 7;

[illegible]

Search completed: September 18, 2002, 16:32:47
Job time: 3847 sec

This Page Blank (uspto)

A:Title: identification and characterisation of a Leishmania donovani antigen belonging
A:Reference number: S11448; MUID:90306037
A:Accession: S11448
A:Molecule type: DNA
A:Residues: 1-653 <MAC1>
A:Cross-references: EMBL:X52314
R:MacFarlane, J.
submitted to the EMBL Data Library, April 1990
A:Reference number: S11572
A:Accession: S11572
A:Molecule type: DNA
A:Residues: 1-208, 'S', 210-653 <MAC2>
A:Cross-references: EMBL:X52314; NID:g9495; PIDN:CAA36551.1; PID:g9496
C:Genetics:
A:Gene: hsc70
C:Function:
A:description: involved in protein folding and assembling/disassembling of protein comp
C:Superfamily: heat shock protein 70
C:Keywords: ATP; molecular chaperone

```

Query Match          9.7%; Score 81; DB 2, Length 653;
Best Local Similarity 23.4%; Pct. No. 6.3;
Matches 36; Conservative 2; Mismatches 53; Indels 42; Gaps

QY 48 LCSSCDLGGENLQLDLPDRCGCCOEAEQFTKKLVAGATILEVCGXKLGKPGFOVAQVRS 107
    | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 451 MTKDCHLLGTFLDLSGIPAPRGVQLETFD---LDANGILNVSAEKSGKRRNQIITN 507

QY 108 DKPKLFR-----GLQ-----IKYVSGSDPYVLKIDDN 134
    | | | | | : | | | | | : | | | | | : | | | | | :
Db 508 DKGRLSKDEIERMVDAMKRYEADRAORDRYEAKNGGLENYASMKNTLGDSSNVSGKLDDS 567

QY 135 --GNIAELSTLTKW---NTDSVEPFLSEKIERI 162
    | | | | | : | | | | | : | | | | | :
Db 568 DKATLNKEIDVTLLELSSNQDTEEEYEHKREL 601

```

RESULT 3

T46574
dnak-type molecular chaperone B1P precursor [similarity] - soybean

N:Alternate names: endoplasmic reticulum HSC70-cognate binding protein

C:Species: Glycine max (soybean)

C:Date: 18-Feb-2000 #sequence_revision 18-Feb-2000 #text_change 11-May-2000

C:Accession: T46574

R:Rifluoredo, J.E.F.; Cascardo, J.M.; Carolino, S.M.B.; Alvini, F.; Fontes, E.P.B. Braz. J. Plant Physiol. 9, 103-110, 1997

A:Title: Water-stress regulation and molecular analysis of the soybean B1P gene family

A:Reference number: 223087

A:Accession: T46574

A:Status: Preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-668 <FIG>

A:Cross-References: EMBL:AF031241; PIDN:AA86942.1

A:Experimental source: strain Roonoke; clone cUFVBIPI; seed

C:Genetics:

A:Gene: B1P

C:Superfamily: heat shock protein 70

C:Keywords: ATP; endoplasmic reticulum; molecular chaperone; stress-induced protein

F:1-27/domain: signal sequence Status predicted <SI>

F:28-668/Product: dnak-type molecular chaperone B1P Status predicted <MT>

```

Query Match Similarity 9.7%; Score 81; DB 2; Length 668:
Best Local Similarity 22.5%; Pred. No. 6.5;
Matches 40; Conservative 26; Mismatches 62; Indels 50; Gaps 7.7;

QY 5 PSCGIVPAFGKRLATVLIQ-----VASFAGFESSEACREIGSSNIDCSDDLQGF 58
      :::::
DB 446 PNTVIPII--KRSQVETTYQDOQTVSIVQFEGERS-----LTKDCRLILGRF 490
      :::::

QY 59 NLLQDDPCRGCCQEMAEQETKKLVAGALLFEGCKKILGRFPVOYAFVSDKRLFLRNGDOI 118
      :::::

```

[illegible]

RESULT 4
S06158
dnaK-type molecular chaperone hsp70 - Trypanosoma cruzi
N/Alternate names: heat shock protein 70
C/Species: Trypanosoma cruzi
C/Date: 22-Jan-1993 #sequence_revision 22-Jan-1993 #text_change 15-Oct-1999
C/Accession: S06158; A28531; S25649; A44979
R/Bedate: C.A.
submitted to the EMBL Data Library, December 1988
A/Reference number: S06158
A/Accession: S06158
A/Molecule type: DNA
A/Residues: 1-680 <BED>
A/Cross-references: GB:X07083; EMBL:X13690; NID:g10621; PIDN:CAA30115.1; PID:g10622
A/Experimental source: Strain Maracai
R/Requena, J.M.; Lopez, M.C.; Jimenez-Ruiz, A.; de la Torre, J.C.; Alonso, C.
Nucleic Acids Res. 16, 1393-1406, 1988
A/Title: A head-to-tail tandem organization of hsp70 genes in Trypanosoma cruzi.
A/Reference number: A28531; MUID:88157703
A/Accession: A28531
A/Molecule type: DNA
A/Residues: 222-473, L', 475-521, 550-622, 'A', 624-630, 'S', 632-680 <RED>
A/Cross-references: GB:X07083; GB:X13690; NID:g10621; PIDN:CAA30115.1; PID:g10622
A/Experimental source: Strain Maracai
A/Note: The sequence is revised in GenBank entry TCHSP70, release 111.0, (PIDN:CAA301
R/Byang, S.; Bergman, L.W.; Scholl, D.R.; Rowland, E.C.
submitted to the EMBL Data Library, August 1992
A/Description: Cloning and characterization of cDNA encoding glucose-regulated protein

A:Molecule type: mRNA
A:Residues: 1-46, 'T', 48-63, 'T', 65, 'SP', 68-167, 'M', 169-178, 'R', 180-240, 'S', 242-256, 'T', 518-522, 'S', 524-529, 'SO', 532-533, 'E', 535, 'RRR', 539-554, 'T', 556-557, 'E', 559-570, 'N', 57
A:Cross-references: EMBL:X67716; NID:910625; PIDN:OAA4952.1; PID:g10626
A:Experimental source: strain Brazil
R:Engman, D.M.; Sias, S.R.; Gabe, J.D.; Doneilson, J.E.; Dragon, E.A.
Mol. Biochem. Parasitol. 37, 285-287, 1989
A:Title: Comparison of HSP70 genes from two strains of Trypanosoma cruzi.
A:Reference number: A44979; MUID:90114336
A:Accession: A44979
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-46, 'T', 48-63, 'T', 65-167, 'M', 169-240, 'S', 242-256, 'T', 258, 'SO', 261-310, 'E', 'R', 312-532, 'S', 534-535, 'RRR', 539-554, 'T', 556-557, 'E', 559-570, 'N', 572-581, 'C', 583-593, 'S', 595-596, 'T', 598-603, 'T', 605-606, 'T', 608-613, 'T', 615-616, 'T', 618-623, 'T', 625-626, 'T', 628-633, 'T', 635-636, 'T', 638-643, 'T', 645-646, 'T', 648-653, 'T', 655-656, 'T', 658-663, 'T', 665-666, 'T', 668-673, 'T', 675-676, 'T', 678-683, 'T', 685-686, 'T', 688-693, 'T', 695-696, 'T', 698-703, 'T', 705-706, 'T', 708-713, 'T', 715-716, 'T', 718-723, 'T', 725-726, 'T', 728-733, 'T', 735-736, 'T', 738-743, 'T', 745-746, 'T', 748-753, 'T', 755-756, 'T', 758-763, 'T', 765-766, 'T', 768-773, 'T', 775-776, 'T', 778-783, 'T', 785-786, 'T', 788-793, 'T', 795-796, 'T', 798-803, 'T', 805-806, 'T', 808-813, 'T', 815-816, 'T', 818-823, 'T', 825-826, 'T', 828-833, 'T', 835-836, 'T', 838-843, 'T', 845-846, 'T', 848-853, 'T', 855-856, 'T', 858-863, 'T', 865-866, 'T', 868-873, 'T', 875-876, 'T', 878-883, 'T', 885-886, 'T', 888-893, 'T', 895-896, 'T', 898-903, 'T', 905-906, 'T', 908-913, 'T', 915-916, 'T', 918-923, 'T', 925-926, 'T', 928-933, 'T', 935-936, 'T', 938-943, 'T', 945-946, 'T', 948-953, 'T', 955-956, 'T', 958-963, 'T', 965-966, 'T', 968-973, 'T', 975-976, 'T', 978-983, 'T', 985-986, 'T', 988-993, 'T', 995-996, 'T', 998-1003, 'T', 1005-1006, 'T', 1008-1013, 'T', 1015-1016, 'T', 1018-1023, 'T', 1025-1026, 'T', 1028-1033, 'T', 1035-1036, 'T', 1038-1043, 'T', 1045-1046, 'T', 1048-1053, 'T', 1055-1056, 'T', 1058-1063, 'T', 1065-1066, 'T', 1068-1073, 'T', 1075-1076, 'T', 1078-1083, 'T', 1085-1086, 'T', 1088-1093, 'T', 1095-1096, 'T', 1098-1103, 'T', 1105-1106, 'T', 1108-1113, 'T', 1115-1116, 'T', 1118-1123, 'T', 1125-1126, 'T', 1128-1133, 'T', 1135-1136, 'T', 1138-1143, 'T', 1145-1146, 'T', 1148-1153, 'T', 1155-1156, 'T', 1158-1163, 'T', 1165-1166, 'T', 1168-1173, 'T', 1175-1176, 'T', 1178-1183, 'T', 1185-1186, 'T', 1188-1193, 'T', 1195-1196, 'T', 1198-1203, 'T', 1205-1206, 'T', 1208-1213, 'T', 1215-1216, 'T', 1218-1223, 'T', 1225-1226, 'T', 1228-1233, 'T', 1235-1236, 'T', 1238-1243, 'T', 1245-1246, 'T', 1248-1253, 'T', 1255-1256, 'T', 1258-1263, 'T', 1265-1266, 'T', 1268-1273, 'T', 1275-1276, 'T', 1278-1283, 'T', 1285-1286, 'T', 1288-1293, 'T', 1295-1296, 'T', 1298-1303, 'T', 1305-1306, 'T', 1308-1313, 'T', 1315-1316, 'T', 1318-1323, 'T', 1325-1326, 'T', 1328-1333, 'T', 1335-1336, 'T', 1338-1343, 'T', 1345-1346, 'T', 1348-1353, 'T', 1355-1356, 'T', 1358-1363, 'T', 1365-1366, 'T', 1368-1373, 'T', 1375-1376, 'T', 1378-1383, 'T', 1385-1386, 'T', 1388-1393, 'T', 1395-1396, 'T', 1398-1403, 'T', 1405-1406, 'T', 1408-1413, 'T', 1415-1416, 'T', 1418-1423, 'T', 1425-1426, 'T', 1428-1433, 'T', 1435-1436, 'T', 1438-1443, 'T', 1445-1446, 'T', 1448-1453, 'T', 1455-1456, 'T', 1458-1463, 'T', 1465-1466, 'T', 1468-1473, 'T', 1475-1476, 'T', 1478-1483, 'T', 1485-1486, 'T', 1488-1493, 'T', 1495-1496, 'T', 1498-1503, 'T', 1505-1506, 'T', 1508-1513, 'T', 1515-1516, 'T', 1518-1523, 'T', 1525-1526, 'T', 1528-1533, 'T', 1535-1536, 'T', 1538-1543, 'T', 1545-1546, 'T', 1548-1553, 'T', 1555-1556, 'T', 1558-1563, 'T', 1565-1566, 'T', 1568-1573, 'T', 1575-1576, 'T', 1578-1583, 'T', 1585-1586, 'T', 1588-1593, 'T', 1595-1596, 'T', 1598-1603, 'T', 1605-1606, 'T', 1608-1613, 'T', 1615-1616, 'T', 1618-1623, 'T', 1625-1626, 'T', 1628-1633, 'T', 1635-1636, 'T', 1638-1643, 'T', 1645-1646, 'T', 1648-1653, 'T', 1655-1656, 'T', 1658-1663, 'T', 1665-1666, 'T', 1668-1673, 'T', 1675-1676, 'T', 1678-1683, 'T', 1685-1686, 'T', 1688-1693, 'T', 1695-1696, 'T', 1698-1703, 'T', 1705-1706, 'T', 1708-1713, 'T', 1715-1716, 'T', 1718-1723, 'T', 1725-1726, 'T', 1728-1733, 'T', 1735-1736, 'T', 1738-1743, 'T', 1745-1746, 'T', 1748-1753, 'T', 1755-1756, 'T', 1758-1763, 'T', 1765-1766, 'T', 1768-1773, 'T', 1775-1776, 'T', 1778-1783, 'T', 1785-1786, 'T', 1788-1793, 'T', 1795-1796, 'T', 1798-1803, 'T', 1805-1806, 'T', 1808-1813, 'T', 1815-1816, 'T', 1818-1823, 'T', 1825-1826, 'T', 1828-1833, 'T', 1835-1836, 'T', 1838-1843, 'T', 1845-1846, 'T', 1848-1853, 'T', 1855-1856, 'T', 1858-1863, 'T', 1865-1866, 'T', 1868-1873, 'T', 1875-1876, 'T', 1878-1883, 'T', 1885-1886, 'T', 1888-1893, 'T', 1895-1896, 'T', 1898-1903, 'T', 1905-1906, 'T', 1908-1913, 'T', 1915-1916, 'T', 1918-1923,

```

Query Match          9.7%; Score 81; DB 2; Length 680;
Best Local Similarity 27.8%; Pred. No. 6.6;
Matches 37; Conservative 16; Mismatches 52; Indels 28; Gaps 4;

QY  48  ICSSDILLGQFLLDLPDPCGCCCEAEQFEKTKLYAGAILLEVCXKIGKRFPOVAFYRS 107
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  452  MKKDCHILCTFELSCIPPRKGVPIQLEYFD---LDANILVNSAEKGTGKRNQIVLTIN 508
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  108  DKPKTIFRGLQIQYVYR-----GSDPVYLKLDDNG-----NIAEEL--S 142
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  509  DKGRISRAEIEEMVEEAKYEAEEDDQVQIQIDAKKGLENYAFSMKNVANDPNVACITEEA 568
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```


QY 143 ILKMTDSEVEEL 155
DB 569 DKRTTSAVEAL 581

RESULT 5

A25398
dnak-type molecular chaperone - Trypanosoma brucei
N:Alternate names: heat shock protein 70
C:Species: Trypanosoma brucei
C>Date: 19-Nov-1988 #sequence_revision 19-Nov-1988 #text_change 13-Mar-1998
C:Accession: A25398
R:Glass, D.J.; Polvere, R.I.; Van der Ploeg, L.H.T.
Mol. Cell. Biol. 6, 4657-4666, 1986
A:title: Conserved sequences and transcription of the hsp70 gene family in Trypanosoma
A:Reference number: A25398; MUID:87089827
A:Accession: A25398
A:Molecule type: DNA
A:Residues: 1-661 <GLA>
A:Cross-References: GB:M4697
C:Function:
A:Description: Involved in protein folding and assembling/disassembling of protein comp
C:Superfamily: heat shock protein 70
C:Keywords: ATP; molecular chaperone

Query Match 9.7%; Score 80.5; DB 2; Length 661;
Best Local Similarity 24.7%; Pred. No. 7.2;
Matches 36; Conservative 16; Mismatches 55; Indels 39; Gaps 4;

QY 48 LCSSCDLQGFNLQDPPCCGCCQGEAOFETKLYAGAILVEYCGKLGFPQVQAFVRS 107
DB 452 MKRDCILGTFDLISGIPAPRGVQIEVTFD---LDANGILSYSAEKGCGKRNQIYITN 508
QY 108 DRPKLFR-----GLQ-----IKYVRGSDPVLKILDDN 134
DB 509 DKGRLSKAIETRMVSPAAYTEADKAHVXXIIDAKNGULENVAFSMKNTINDPNVAGKILDDA 568
QY 135 GNIAEELSI--LKMTDSEVEELSE 157
DB 569 DKNAVTTAVEALRWLNDQNEASLEE 594

RESULT 6

C87482
this protein [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C:Accession: C87482
R:Nierman, W.C.; Feldblum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Emla, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: C87482
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-269 <STO>
A:Cross-References: GB:AE005673; NID:g13423327; PIDN:AAK23855.1; GSPDB:GN00148
C:Genetics:
A:Gene: CCI880
C:Superfamily: thiamin biosynthesis protein thiG

Query Match 9.6%; Score 80; DB 2; Length 269;
Best Local Similarity 29.4%; Pred. No. 2.9;
Matches 42; Conservative 15; Mismatches 66; Indels 20; Gaps 5;

QY 9 LVPAGKRLILATVLAQVAFGAFFSSEACRELGFSNNLLCSCDLL-----GQFNLLQ 62
DB 29 LTVGCKRYDYATNMAAAAGAGETVAVRRV-----NLTPSQPLLVIVYKFTERTYL- 83

QY 63 LDPDCRCQGEAOFETKLYAGA-----ILEYCGKLGKRGFPQVQAFVRSNDRKPLRGL 116
DB 84 --PNTAGCFTGEDAVRTLRAREAGMDLVKLEVLSDPKTLTFPDMETLSIKLLVADGF 141

QY 117 QIKYVRGSDPV-LKILDDNGNIA 138
DB 142 QVMVYCSDDPVYARKLEEAGAVA 164

RESULT 7

S21877
dnak-type molecular chaperone b1p1 - common tobacco (fragment)
N:Alternate names: luminal binding protein b1p1
C:Species: Nicotiana tabacum (common tobacco)
C>Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 20-Aug-1999
C:Accession: S21877; PQ0264
R:Denecke, J.; Goldman, M.H.; Demolder, J.; Seurinck, J.; Botterman, J.
submitted to the EMBL Data Library, June 1991
A:Description: The luminal binding protein (B1p) is encoded by a multigene family in
A:Reference number: S21877
A:Accession: S21877
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-290 <DEN>
A:Cross-References: EMBL:X60060; NID:g19804; PIDN:CAA42662.1; PID:g19805
R:Denecke, J.; Goldman, M.H.S.; Demolder, J.; Seurinck, J.; Botterman, J.
Plant Cell 3, 1025-1035, 1991
A:title: The tobacco luminal binding protein is encoded by a multigene family.
A:Reference number: JQ1360; MUID:92361242
A:Accession: PQ0264
A:Molecule type: mRNA
A:Residues: 263-290 <DEW>
A:Note: translation of the nucleotide sequence is not complete
C:Genetics:
A:Gene: b1p1
C:Function:
A:Description: Involved in protein folding and assembling/disassembling of protein co
uration steps in vivo
C:Superfamily: heat shock protein 70
C:Keywords: ATP; endoplasmic reticulum; molecular chaperone
F;287-290/Region: endoplasmic reticulum retention signal

Query Match 9.4%; Score 78; DB 2; Length 290;
Best Local Similarity 23.0%; Pred. No. 5;
Matches 41; Conservative 24; Mismatches 63; Indels 50; Gaps 7;

QY 5 PSCCLVPAGCKRLLLATVLA-----AVSAFGAFSSSEACRELGFSNNLLCSCDLLGQF 58
DB 71 PRNTVPIPT--KKSQVFTTYQDOOTVSIQVEGERS-----LTKDCRLLGKF 115
QY 59 NLQLDPPCCGCCQGEAOFETKLYAGAILVEYCGKLGKRGFPQVQAFVRSRKLFRGLQI 118
DB 116 DLGVIAPAPRGVQIEVTFD--ANGILNVAEDKGGCKSKRTITNDKGR----- 166
QY 119 KYVRGSDPVLKILDDNGNIAEELSIKMTND---SVEEF-----ISEKLE 160
DB 167 ----SQEIERMVRAEAEFEEDKKYKERIDANSLETTYNNKKNQINDKRLADKLE 220

RESULT 8

S21878
dnak-type molecular chaperone b1p2 - common tobacco (fragment)
N:Alternate names: luminal binding protein b1p2
C:Species: Nicotiana tabacum (common tobacco)
C>Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 20-Aug-1999
C:Accession: PQ0262; S21878
R:Denecke, J.; Goldman, M.H.S.; Demolder, J.; Seurinck, J.; Botterman, J.
Plant Cell 3, 1025-1035, 1991
A:title: The tobacco luminal binding protein is encoded by a multigene family.
A:Reference number: JQ1360; MUID:92361242
A:Accession: PQ0262
A:Molecule type: mRNA

QY 108 DKPKLFRGLQIRYVNGSDPVKLKLD 133
 Db 508 DKGRL-----SKDEIERMVD 523

RESULT 12

T06598
 dnak-type molecular chaperone Bip-A - soybean
 N:Alternate names: binding protein Bip isoform A
 C:Species: Glycine max (soybean)
 C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
 C:Accession: T06598
 R:Kalinowski, A.; Kowley, D.L.; Loer, D.S.; Foley, C.; Buta, G.; Herman, E.M.
 A:Title: Binding-protein expression is subject to temporal, developmental and stress-ind
 A:Reference number: 215624; MUID:95218610
 A:Accession: T06598
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-664 <KAL>
 A:Cross-references: EMBL:U08384; NID:q475601; PIDN:AAA81956.1; PID:q475602
 A:Experimental source: strain Century 84; leaf
 C:Function:
 A:Description: involved in protein folding and assembling/disassembling of protein comp
 C:Superfamily: heat shock protein 70
 C:Keywords: ATP; molecular chaperone

Query Match 9.2%; Score 76.5; DB 2; Length 664;

Best Local Similarity 22.2%; Pred. No. 18; Mismatches 45; Conservative 26; Mismatches 69; Indels 63; Gaps 8;

QY 5 PSCGLVPAGKRLATVLYQ-----AVSAGFAEFSSEACRELGFSSNLLCSSCDLLGQF 58
 Db 444 PRNTVIPT--KKSQVFTTYQDOQTVTISVFEGERS-----LTFDCRLGKF 488
 QY 59 NLQLDPCRCGCCOEAPETKRLVAGALVEYCGKLLGFPOVOAFVRSDKRLFRGLQI 118
 Db 489 ELSGPPAPRGTPQILEVFEVD--ANGILNKAKEDKGGKSEKITITNEKGRLSQEEIE 545
 QY 119 KYVNGSD-----PVLKLD-----DNGNIAREL----- 141
 Db 546 RMVREAEFEAEEDKKVKRIRIDARNSLETYYVMKNQVSDKDLKLSDEKEKVEYAVK 605
 QY 142 SILKMTD--SVEEFLSEKLERI 162
 Db 606 EALEWLDNDNGSVKRYEKKKEV 628

RESULT 13

S77181
 hypothetical protein slr1809 - Synechocystis sp. (strain PCC 6803)
 C:Species: Synechocystis sp.
 A:Variety: PCC 6803
 C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
 C:Accession: S77181
 R:Kaneke, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
 O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
 DNA Res. 3, 109-136, 1996
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
 S.
 A:Reference number: S74322; MUID:97061201
 A:Accession: S77181
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-265 <KAN>
 A:Cross-references: EMBL:D90908; GB:AA001339; NID:q1652725; PIDN:BA117739.1; PID:0101847
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 9.1%; Score 76; DB 2; Length 265;
 Best Local Similarity 23.2%; Pred. No. 7.1;

Matches 41; Conservative 22; Mismatches 60; Indels 54; Gaps 8;

QY 31 AEFSSAEARELAFSSNLLCSSCDLLGQF-----NLQLDPCRCGCCOE 73
 Db 74 AELDLDAIAFNPPESKRLSSAEAYLGKMKQHRRAPADQMHLRVRCLELPCCBCLADL 133
 QY 74 EAOFETK-----LVGAILVECGKXL-----GRPOVOAFVRSDKRLFRGLQ 117
 Db 134 LAEFRRCQRYRFTYISLMFAGAIL--CGGGIYLLSGCLP---FVIEKART-ENIQ 186
 QY 118 IYVNGSDPVKLKLDNGNIARELSTL-----KMTDSVEEF-----LSEKLE 160
 Db 187 STIEAMNDITADLMGEQALRREPDTLISQQLQENLNDRLENLQARLKALEEKIK 243

RESULT 14

S21880
 dnak-type molecular chaperone b1p5 precursor - common tobacco
 N:Alternate names: luminal binding protein b1p5
 C:Species: Nicotiana tabacum (common tobacco)
 C:Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 20-Aug-1999
 C:Accession: S21880; J01361
 R:Denecke, J.; Goldman, M.H.; Demolder, J.; Seurinck, J.; Botterman, J.
 Submitted to the EMBL Data Library, June 1991
 A:Description: The luminal binding protein (b1p) is encoded by a multigene family in
 A:Reference number: S21877
 A:Accession: S21880
 A:Molecule type: mRNA
 A:Residues: 1-668 <DEN>
 A:Cross-references: EMBL:X60058; NID:q19812; PIDN:CAA42660.1; PID:q19813
 R:Denecke, J.; Goldman, M.H.S.; Demolder, J.; Seurinck, J.; Botterman, J.
 Plant Cell 3, 1025-1035, 1991
 A:Title: The tobacco luminal binding protein is encoded by a multigene family.
 A:Reference number: J01360; MUID:92361242
 A:Accession: J01361
 A:Molecule type: mRNA
 A:Residues: 1-653, 'G', 655-668 <DEN>
 A:Note: translation of the nucleotide sequence is not complete
 C:Genetics:
 A:Gene: b1p5
 C:Function:
 A:Description: involved in protein folding and assembling/disassembling of protein co
 uration steps in vivo
 C:Superfamily: heat shock protein 70
 C:Keywords: ATP; endoplasmic reticulum; molecular chaperone
 F.1-24/Domain: signal sequence #status predicted <SIG>
 F.665-668/Region: endoplasmic reticulum retention signal

Query Match 9.1%; Score 76; DB 2; Length 668;

Best Local Similarity 23.0%; Pred. No. 20; Mismatches 41; Conservative 23; Mismatches 64; Indels 50; Gaps 7;

QY 5 PSCGLVPAGKRLATVLYQ-----AVSAGFAEFSSEACRELGFSSNLLCSSCDLLGQF 58
 Db 447 PRNTVIPT--KKSQVFTTYQDOQTVTISVFEGERS-----LTFDCRLGKF 491
 QY 59 NLQLDPCRCGCCOEAPETKRLVAGALVEYCGKLLGFPOVOAFVRSDKRLFRGLQI 118
 Db 492 DLTGTPAPRGTPQILEVFEVD--ANGILNKAKEDKAGSKSEKITITINDKRL----- 542
 QY 119 KYVNGSDPVKLKLDNGNIARELSTLKNMTD--SVEEF-----LSEKLE 160
 Db 543 -----SQEEIERMVKKEAEFEAEEDKKVKRIRIDARNSLETYYVMKNQVSDKDLKLSDEKEKVEYAVK 596

RESULT 15

S14875
 dnak-type molecular chaperone hsp70 - Trypanosoma cruzi (fragment)
 N:Alternate names: heat shock protein 70
 C:Species: Trypanosoma cruzi
 C:Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 20-Aug-1999
 C:Accession: S14875

R;Levy Yeyati, P.; Lafon, S.; Bonnefoy, S.; Levi, M., submitted to the EMRL Data Library April 1991

submitted to the EMBL Data Library, April 1991

A; Description: Nucleotide sequence of a Trypanosoma cruzi HSP-70 c-DNA.

A:Reference number: S14875

A;Accession: S14875

A; molecule type: mRNA

A; Residues: 1-261 <LEV>

A; Cross-references: EMBL:X58715; NID:g10623; PIDN:CAA1551.1; PID:g10624

C; Genetics:

C;GENETICS:
A;Gene: hsp70

A; Gene: *usp*
C; Function:

A;Description: Involved in protein folding and assembling/disassembling of protein complex

C;Superfamily: heat shock protein 70

c;superfamily: heat shock protein 70
c;keywords: ATP; molecular chaperone

Query Match	Score	DB 2;	Length
9.18;	75.5;	261;	
33.98;	7.8;		

Best Local Similarity 33.8%; Pred. No. 7.8;
Matches 33. Concomitance 0. Mismatches 31. Tieds 3. Cases 1.

Matches 22; Conservative 9; Mismatches 31; Indels 3; Gaps 1;

48 LCSSCDLGFNLLQDDPCRCGCCQEEAQFETKTLVAGALEVCGKKLGRFPQVQAFVRS 107

```
Db 52 MTKDCHLGTDLGIPAPRGVPQIEVTFD--LDANGILNVSAAEKGTKRNGIYITN 108
```

OY 108 DKPKL 112

11 : 1
109 PKGBT. 113

Search completed: September 18, 2002, 16:33:51

Job time: 3691 sec